

Where We Come from, Where We Go – Describing Population Change in New Zealand

Keith McLeod

Analytical Paper 18/02

April 2018

DISCLAIMER

The views, opinions, findings, and conclusions or recommendations expressed in this report are strictly those of the authors. They do not necessarily reflect the views of the New Zealand Treasury, Statistics New Zealand or the New Zealand Government. The New Zealand Treasury and the New Zealand Government take no responsibility for any errors or omissions in, or for the correctness of, the information contained in this Analytical Paper.

The results in this report are not official statistics, they have been created for research purposes from the Integrated Data Infrastructure (IDI) managed by Statistics New Zealand.

Access to the anonymised data used in this study was provided by Statistics NZ in accordance with security and confidentiality provisions of the Statistics Act 1975. Only people authorised by the Statistics Act 1975 are allowed to see data about a particular person, household, business or organisation and the results in this paper have been confidentialised to protect these groups from identification.

Careful consideration has been given to the privacy, security and confidentiality issues associated with using administrative and survey data in the IDI. Further detail can be found in the Privacy impact assessment for the Integrated Data Infrastructure available from www.stats.govt.nz.

The results are based in part on tax data supplied by Inland Revenue to Statistics NZ under the Tax Administration Act 1994. This tax data must be used only for statistical purposes, and no individual information may be published or disclosed in any other form, or provided to Inland Revenue for administrative or regulatory purposes.

Any person who has had access to the unit-record data has certified that they have been shown, have read, and have understood section 81 of the Tax Administration Act 1994, which relates to secrecy. Any discussion of data limitations or weaknesses is in the context of using the IDI for statistical purposes, and is not related to the data's ability to support Inland Revenue's core operational requirements.

ANALYTICAL PAPER 18/02 | Where We Come from, Where We Go – Describing Population Change in New Zealand

MONTH/YEAR | April 2018

AUTHOR | Keith McLeod

ISBN (ONLINE) | 978-1-98-853498-5

URL | Treasury website at April 2018:
<https://treasury.govt.nz/publications/ap/ap-18-02>
Persistent URL: <http://purl.oclc.org/nzt/p-2029>

ACKNOWLEDGEMENTS | The author is grateful to Kim Dunstan, Kirsten Nissen, and Nathaniel Matheson-Dunning from Stats NZ, and Mary Adams from the Ministry of Business, Innovation, and Employment for their extremely useful comments on drafts of this paper.

The author is also grateful to Danny Wu, Sam Harvey, Shirley Wu, and Phil Shepherd from Harmonic Analytics Limited for their work on developing visualisations for the Insights tool which sits alongside this report.

Finally the author thanks Tod Wright, Meghan Stephens, and Robert Templeton from The Treasury for their technical advice on the approach taken, and helpful comments on this paper.

Any remaining errors are the sole responsibility of the author.

CODE AVAILABILITY | The code used to produce the statistics used in this report can be accessed at the following GitHub address:
<https://github.com/Treasury-Analytics-and-Insights/>

NZ TREASURY | New Zealand Treasury
PO Box 3724
Wellington 6008
NEW ZEALAND

Email | information@treasury.govt.nz
Telephone | 64-4-472 2733
Website | www.treasury.govt.nz

PURPOSE OF THE ANALYTICAL PAPERS SERIES | The Treasury's aim in publishing the Analytical Papers series is to make this analysis available to a wider audience and to inform and encourage public debate, with the ultimate aim of informing our policy advice.

Analytical Papers are commissioned as part of the Treasury's core function in developing and providing advice to Ministers. They include work undertaken by staff at the Treasury or other government departments, as well as work undertaken for the Treasury by external researchers or consultants.

Analytical Papers do not themselves represent policy advice.

© Crown Copyright



This work is licensed under the Creative Commons Attribution 4.0 International licence. In essence, you are free to copy, distribute and adapt the work, as long as you attribute the work to the Crown and abide by the other licence terms.

To view a copy of this licence, visit <https://creativecommons.org/licenses/by/4.0/>. Please note that no departmental or governmental emblem, logo or Coat of Arms may be used in any way that infringes any provision of the [Flags, Emblems, and Names Protection Act 1981](#). Attribution to the Crown should be in written form and not by reproduction of any such emblem, logo or Coat of Arms.

Executive summary

This paper describes new sub-national New Zealand population measures that have been developed by The Treasury using integrated administrative data. The new measures have been made available in an interactive online form in the Treasury's Insights tool.¹ The paper describes the development of the new measures, compares them against existing sources of official population data, describes some high level findings, and presents three case studies describing population change in specific territorial authority areas.

The results outlined in this paper highlight the enormous potential of integrated administrative data to better understand population change at a sub-national level. New Zealand has a robust system of population estimates, and the data described in this paper has the potential to complement this system. Nevertheless, the results are exploratory in nature, and further work is required to better understand the strengths and limitations of the data. The findings are not official statistics and should be treated with caution.

A particular strength of the analysis outlined here is the ability to measure and describe patterns of internal migration within New Zealand, something that has previously been largely reliant on the 5-yearly census. The analysis not only describes patterns of internal migration, but sets these alongside other key dimensions of population change: ageing, natural increase, and international migration.

Overall, the data described in the paper appears to be of high quality, matching well at both an individual and aggregate level with other sources of population change information. Although individual-level internal migration is not precisely identified, these errors do not seem to have a large impact at an aggregate level, and estimates are not subject to the same level of non-response bias that affects Census-based estimates.

Because the approach outlined in this paper is reliant on the collection of administrative data, and its subsequent incorporation in the Integrated Data Infrastructure (IDI), estimates are not immediately available for release. Results for a particular calendar year are only likely to be able to be produced nine or more months after the end of that year. As such, results for 2017 are only likely to be able to be produced in late 2018.

Through a better understanding of territorial authority area population change, local and central government decision-makers, businesses, and the general public, will be better able to understand the changing needs of different communities. The new data also opens up possibilities for researchers to explore the impacts of such change on outcomes of interest to people and communities. The three case studies provide a small taste of this potential:

- The data shows the effects of two large-scale earthquakes on the Christchurch population. The population fell by almost 20,000 over a two-year period, but has rebounded in the subsequent three years. The initial drop was largely fuelled by internal migration to many other areas of New Zealand, while the rebounding population has been largely fuelled by international migration, especially from India and the Philippines. In recent years internal migration out of Christchurch City has

¹ Insights, including the new Population explorer and Population directions tools, is accessible at <https://insights.apps.treasury.govt.nz/>.

been largely to the adjacent Selwyn and Waimakariri areas, contributing to large-scale growth in the Greater Christchurch area population since 2012.

- The population of the tourism-focussed Queenstown-Lakes district increased by around 10,000 people, or 40 percent, between 2008 and 2016. In recent years this growth has accelerated, with a growth of almost 2,500 people in 2016 alone. Migration from other areas of New Zealand has contributed to this growth, but the acceleration has been largely driven by migration from overseas, especially of temporary work visa holders. The United Kingdom has been a key source of growth in recent years, but most recently a number of other countries, such as Brazil and Australia, have also been key contributors.
- Auckland, New Zealand's largest city, has also experienced year-on-year growth since 2008. This has been driven largely by migration from overseas, with foreign migrants more than offsetting net losses of New Zealanders moving away. Since 2012 increasing numbers of people have been leaving Auckland to move to other areas, especially Tauranga, Waikato District, Whangarei, and the Far North. This has slowed population increase in Auckland over that period.

Although the case studies presented here tell a similar story to official population estimates, there are some differences, particularly in Auckland, where our estimates show much lower population growth in recent years. Our estimate of Auckland population growth due to net migration between 2013 and 2016 is about half the official figure. More work is required to better understand these differences. The difference could derive from the difficulty in determining people's location of residence after their arrival in New Zealand in either or both of the sources, or may relate to the different residence definitions adopted.

The Insights Population explorer and Population directions tools enable users to generate their own insights from the data produced in this study, exploring the areas and aspects of change they are most interested in. The code used to construct the data described in this paper has been shared online, enabling researchers to further develop and improve the data, generate new descriptive data from other administrative and survey sources, and undertake research to better understand the changing population of New Zealand and its communities.

Contents

Executive summary	i
1 Introduction	1
2 Constructing the data	4
2.1 Constructing the study population	4
2.2 Sources of location data in the IDI.....	6
2.3 Assigning locations to the study population.....	9
2.4 Decomposing and describing population change	16
2.5 Quality assessment.....	21
3 Key results	32
3.1 High level results.....	32
3.2 Case study 1 – Christchurch.....	37
3.3 Case study 2 – Queenstown-Lakes District.....	39
3.4 Case study 3 – Auckland	41
4 Next steps	43
References	44
Appendix	45

List of figures

Figure 1: Age at which location records are collected, by record source, health-related sources, year to June 2016.....	7
Figure 2: Age at which location records are collected, by record source, education-related sources, year to June 2016.....	7
Figure 3: Age at which location records are collected, by record source; employment and welfare sources, and motor vehicle licences, year to June 2016	8
Figure 4: Age at which location records are collected, by record source, other sources, year to June 2016	8
Figure 5: Accuracy of TA location by source and years since data was collected	11
Figure 6: Comparison of different estimated resident populations over time.....	14
Figure 7: Comparison of different 2013 resident populations by single year of age to age 90	14
Figure 8: Percentage of IDI ERP unable to be allocated to a TA, 2008 to 2016	15
Figure 9: Residence classification example, 12/16 rule.....	16
Figure 10: Internal migration flow comparison of Census and IDI-based figures – All combinations of TAs 2008 to 2013	27
Figure 11: Net internal migration flow comparison of Census and IDI-based figures – All combinations of TAs 2008 to 2013	28
Figure 12: Net internal migration flow comparison of Census and IDI-based figures – All TAs 2008 to 2013	29
Figure 13: International migration comparison of official and IDI-based figures	30
Figure 14: Natural increase comparison of official and IDI-based figures	31
Figure 15: Total NZ decomposition of annual net population change, 2008 to 2016	32
Figure 16: Internal migration rate, 2008 to 2016.....	33
Figure 17: Internal migration rate by country of origin, 2008 to 2016	33
Figure 18: Total NZ net population change due to international migration by country of origin, 2008 to 2016 calendar years	34
Figure 19: Total NZ net population change due to international migration by visa type, 2008 to 2016 calendar years.....	35

List of tables

Table 1: Country and region of origin classification	20
Table 2: Use and quality of location records by data source	21
Table 3: Use and quality of location records by time since record collected	22
Table 4: Location assignment and quality by sex and age	23
Table 5: Location assignment and quality of 2008 location assignment.....	24
Table 6: Location assignment and quality of HLFs/HES respondents over time	25
Table 7: Matching of Census movers and IDI movers	25
Table 8: Matching of 2008 and 2013 TA for Census movers and non-movers	26
Table 9: Match to 2013 location for Census movers and non-movers.....	26
Table 10: Sources of location data in the IDI	45

Where We Come from, Where We Go – Describing Population Change in New Zealand

1 Introduction

This paper describes new sub-national New Zealand population measures that have been developed by The Treasury using integrated administrative data. The paper accompanies the release of interactive online tools that present new data about the changing population of New Zealand and its territorial authority (TA) areas. The Insights Population explorer and Population directions tools can be accessed at <https://insights.apps.treasury.govt.nz/>. The paper provides a technical description of the construction of the data that underpins the tool, a comparison of the data against other sources of similar data, and some high level findings.

Changing populations have an impact on people's lives in a number of different ways. Local authorities and central government must plan for and adapt to the associated changing needs of local populations. Growing populations put additional demand on existing infrastructure, while shrinking or ageing populations raise different challenges. The changing demographic characteristics of populations also place new demands on local area populations and government. For example, people's service needs change considerably across their life course, and an older population will have quite different needs to a younger one.

In order to understand the impact population change has at a subnational level, detailed information is required to better understand what has changed and how it has changed. Official population estimates provide accurate, timely and cohesive insights into population change. However, there is potential for integrated administrative data to provide further insights, particularly into subnational migration flows, whether these are international or within New Zealand. This paper explores the potential for integrated administrative data to be used to fill some of these gaps.

Administrative data may also provide the possibility for: more frequent updating of some population estimates, population estimates by ethnicity at a subnational level outside of Census years, and the use of alternative population definitions. Integrated administrative sources are not subject to some of the sources of error typically encountered in survey-based data collection, such as non-response bias, recall error, and sampling error, but are typically subject to other types of error such as linking error, and the populations captured by such sources may not provide a complete enumeration of the population of interest.

A five-yearly census is undertaken by Statistics New Zealand describing the population at the time the census is conducted. Population estimates based on the census include adjustments for net census under-count (under-count minus over-count) and residents

temporarily overseas, to provide an estimate of the resident population for any given geographic area. That estimate is then updated over time to account for births and deaths, immigration and emigration.

Birth and death data are derived from birth and death registrations, which are highly accurate, while immigration and emigration is based on passport information, departure cards, and arrival cards. People record their intentions in these cards, and this information is used to classify people as 'short-term' or into the much smaller group of 'permanent and long-term' travellers.² The latter group forms the basis for migration statistics.

Internal migration between areas of New Zealand is only directly measured every five years using the Census, with respondents asked for the address they were residing at five years earlier. This limits the amount of information available about internal migration over shorter time periods and since the latest Census. In between Censuses Stats NZ updates local area resident population estimates to account for the effects of internal migration. Since 2012 this has primarily been based on primary health organisation (PHO) enrolment data and Inland Revenue (IR) tax data.

As the different components of population change are measured using different sources at different times in the official statistical system, getting a good understanding of the drivers of local area population change, especially between Censuses, can be difficult. Recent advancements in integrated administrative data, particularly with the development of the Integrated Data Infrastructure (IDI), present an opportunity to construct new measures of population change, and the components of that change. These measures have the potential over time to provide a more complete, consistent, and regularly measured picture of population change.

Stats NZ have sought to utilise these data sources in the construction of new experimental population estimates, assessing the feasibility of alternative sources (Stats NZ 2011), constructing new measures of external migration (Stats NZ 2017a), and most recently constructing new subnational population estimates (Stats NZ 2017b).

This paper takes this work a step further, developing population estimates from the IDI in a similar way to the Stats NZ approach, but then decomposing changes in the population according to whether they are due to internal migration (between areas of New Zealand), external migration (to and from New Zealand), or natural increase (births and deaths). These decomposed population flows are then further described according to various population characteristics, as outlined in Section 2.2. Some high level results are discussed in this paper, while more detailed data visualisations are being made available through The Treasury's Insights tool³.

The current data and analysis covers all but one of the 67 New Zealand territorial authority areas⁴, the equivalent of a city or provincial district, and covers the period from 2008 to 2016. Data to construct estimates for 2017 will become available later in 2018. Future work is likely to look at different levels of geographical disaggregation, particularly in the large Auckland area.

² More recently new measures of migration have been developed that are based on people's actual movements, rather than their intentions. See Stats NZ (2017a) for more information.

³ See <https://insights.apps.treasury.govt.nz/>. Insights consists of a range of online interactive analytical tools. Its original development is discussed in McLeod and Tumen (2017).

⁴ The Chatham Islands TA is excluded due to quality issues, leaving 66 TAs.

This analysis is intended to give policy-makers, planners, and the general population additional information about the changing New Zealand population. By providing information on all aspects of population change it will also enable new research to be undertaken that improves our understanding of the impacts of this change. Finally, as measures of population change are calculated based on individual IDI records, there is considerable potential to supplement these estimates with additional information on individuals – for example, their employment experience, access to healthcare, and use of government services.

2 Constructing the data

Constructing the data for this study involves a number of stages, as outlined in this section.

2.1 Constructing the study population

Standard statistical definitions of residence

Official statistical data sources typically define the population of interest according to a 'usual residence' definition. While some guidance is typically given to people responding to Stats NZ surveys, the concept is regarded as 'self-defined because it involves feelings of belonging, association and participation in and with a household'.⁵ As such, time criteria are typically not applied, with the exception being a 12 month period for determining whether a person resides in New Zealand or not. Even this is only vaguely defined however, leaving much of the determination to individual respondents.

In the 2013 Census respondents were asked where they 'usually live', with the guidance indicating that 'overseas residents' should give a New Zealand address if they 'will be staying in New Zealand for less than 12 months'.⁶ No further guidance is given as to how respondents should decide whether they are 'overseas residents' or not, whether they should include time already spent in the country as part of the 12 month period, or how they should treat short-term absences.

Official migration data also uses a 12 month time period to determine whether somebody is considered to be a 'permanent or long-term' (PLT) migrant. Arriving passengers who 'live in New Zealand' are asked whether they will 'mostly live' in New Zealand for the next 12 months, while other arriving passengers are asked how long they intend to stay in New Zealand⁷. Departure cards take a slightly different approach, with all departing passengers being asked whether they have been 'living, working, or studying in New Zealand for 12 months or more'.⁸

As with usual residence definitions little guidance is given as to how respondents should treat periods of time spent outside of New Zealand. In addition, arriving and departing passengers are defined as being PLT or not largely according to their stated intentions. People may have incentives to not report their intentions accurately (for example, if they could be seen as breaching the conditions of their visa), or their intentions may change, and this impacts on the accuracy of PLT migration estimates.

In birth registration statistics, the residence of the child is based on the self-identified 'home address' of the mother. In death registration statistics, the residence of the deceased is based on their 'usual home address' as identified by the family and/or funeral

⁵ <http://archive.stats.govt.nz/methods/classifications-and-standards/classification-related-stats-standards/usual-residence/definition.aspx>.

⁶ <http://archive.stats.govt.nz/~media/Statistics/Census/2013%20Census/forms/2013-guide-notes.pdf>.

⁷ <http://www.mbie.govt.nz/info-services/sectors-industries/tourism/tourism-research-data/international-travel/documents-image-library/visitor-arrival-card.pdf>.

⁸ <http://www.mbie.govt.nz/info-services/sectors-industries/tourism/tourism-research-data/international-travel/documents-image-library/nz-departure-card.pdf>.

director. It is unclear how the interpretation of these questions aligns with other residence definitions.

Defining usual residence using administrative data

In collecting administrative data, government agencies do not typically ask people whether they are resident in New Zealand. Given that the concept of residence is to a large degree self-determined, this presents a challenge in selecting a resident population using this data. Stats NZ have in recent years developed rules for defining a resident population from the IDI and for identifying PLT migrants using variants of what is called the 12/16 rule (see Stats NZ 2017a and Stats NZ 2017b). We adopt a similar rule to define an IDI-based estimated resident population (IDI ERP) for this study.

The 12/16 rule as applied in both recent Stats NZ papers assesses time in NZ from the point a movement occurs forwards. If a person spends 12 or more of the 16 months following the date of a departure from New Zealand, that departure is considered to be permanent or long-term, and they are removed from the population. Conversely, if a person who is not in the ERP spends 12 months or more in New Zealand in the 16 months following an arrival they are added to the ERP.

We instead centre the 16 month period on the middle of the year of interest. The adopted approach has an advantage in that the resident population can be defined in a slightly more timely fashion⁹. A person is considered to be in our IDI ERP in a particular year if they are in the country for at least 12 months of the 16 month period from November the previous year to February the next. To distinguish the two approaches in this report we refer to our population as the IDI ERP, and Stats NZ's population (as defined in Stats NZ 2017b) as the SNZ IDI ERP.

As with the SNZ IDI ERP we also apply a range of activity measures to confirm that a person is in the population. This is particularly necessary due to migration data not being available in the IDI prior to 1997. If a person were born in New Zealand and left the country prior to 1997 there is no way of knowing of their departure. Without the activity measures they would be assumed to still be in the country, and would be included in the ERP. For the purposes of this study we apply the activity rules over a broader period than has been applied by Stats NZ; a period covering the entire study period of interest, from the 2008 calendar year to the 2016 calendar year. This has the advantage that it removes the flow of people into and out of the ERP that cannot be attributed to international travel, birth or death. It also reduces the risk of under-counting, where someone may be legitimately resident, but not counted due to not having any activity, or where that activity is missed due to matching error.

This extended activity period does come with a heightened risk of over-counting the population however. Where a person is duplicated in the data through a matching error, we are more likely to include both records in the ERP with our expanded activity period. In addition, if someone departs New Zealand during the period, but we miss the departure due to matching error, they may be erroneously retained in the ERP. The risk of over-counting should be reduced in future refreshes of the IDI, as Stats NZ have plans to improve the handling of duplicate records (see Stats NZ 2017b for more information).

⁹ Only movements for the first two months of the following year are needed to define the resident population in any calendar year. The standard SNZ approach requires a 16 month window following the date of interest, as departures or arrivals up to that date can only be classified at that point.

2.2 Sources of location data in the IDI

The IDI consists of a wide range of datasets, most of which are collected by government agencies as part of their administrative activities: administering the tax, school, health, and social welfare systems for example. Many of these administrative activities require agencies to collect addresses from people, and these addresses then allow people to be assigned to a geographic location of residence at a particular point in time.

The Appendix provides information about the sources of location data identified for use in this study. Table 10 summarises the key characteristics of these sources, including the number of location records collected, and the period over which they are collected. Sources vary widely according to the period across which data is collected, the demographic characteristics of the population covered, and the quality and timeliness of the location information collected. Some sources record a residential address, while for some other sources we can infer a residential location from other information e.g. the location of the school or tertiary institution they are enrolled in, or the location of the employer they work for.

Location information is also collected in three official statistical survey sources (HES, HLFS and Census), which are subsequently linked to the IDI at an individual level. Location information is not collected consistently by agencies across a person's life, and as such, it's important when using administrative location sources that a number of sources are used to present a complete picture. Figures 1 to 4 show the main sources of location information, and the number of records that are collected in these sources at any particular year of age. These are contrasted against the 2016 IDI ERP to give an idea of the proportion of people for whom different data is collected at different ages.¹⁰

Figure 1 shows health-related sources of information by age, with data tending to be collected at birth and then fairly regularly across all ages. Figure 2 is focussed on education-related sources, with collection focussed on younger ages. Finally Figure 3 and Figure 4 illustrate employment and welfare-related sources, and miscellaneous sources, with information tending to be collected across people's adult years.

A large number of location records are collected in people's first year of life, albeit predominantly from three sources: the National Health Index (NHI) register, the Primary Health Organisation (PHO) register, and Inland Revenue. Over the childhood years many records are also collected from the Accident Compensation Corporation (ACC), but school enrolments become a key source of location information, particularly at ages 5, 11 and 13.

Large number of records are then collected in people's late teens and 20s, possibly due to people's high mobility at these ages, but also due to the fact that many people interact with government agencies for the first time – particularly with Inland Revenue, the Ministry of Social Development, the New Zealand Transport Agency (drivers' and motor vehicle licences), and tertiary providers.

Most sources of addresses collect data less frequently at progressively older ages, however there are further peaks in location record collection for NHI at age 45 (possibly coinciding with the age of eligibility for breast screening), and the Ministry of Social Development (MSD) at age 65, coinciding with the age of eligibility for New Zealand

¹⁰ Note that the number of location records can exceed the IDI ERP for two reasons. Multiple location records can be collected for the same person from the same source in the same year. More importantly, location records may be collected for people who do not meet the residency rules and are not included in the IDI ERP.

Superannuation. This decline in address collection both relates to increasing stability of location, and, at the older ages, to reduction in the population size as people die.

Figure 1: Age at which location records are collected, by record source, health-related sources, year to June 2016

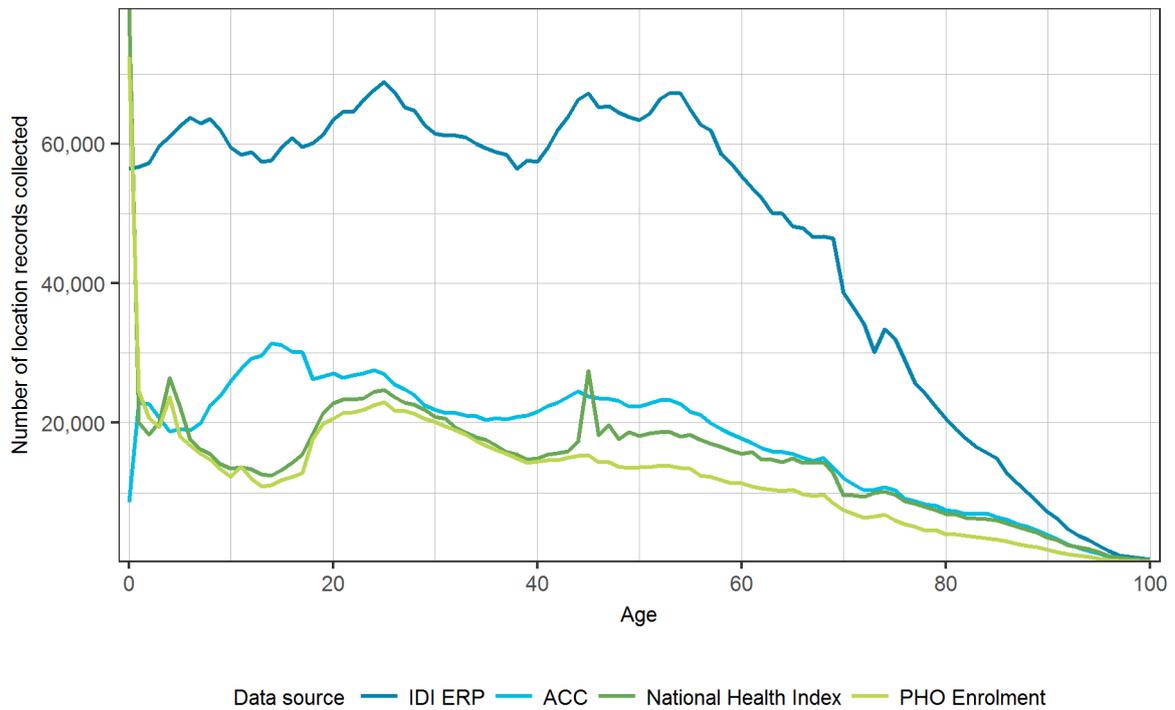


Figure 2: Age at which location records are collected, by record source, education-related sources, year to June 2016

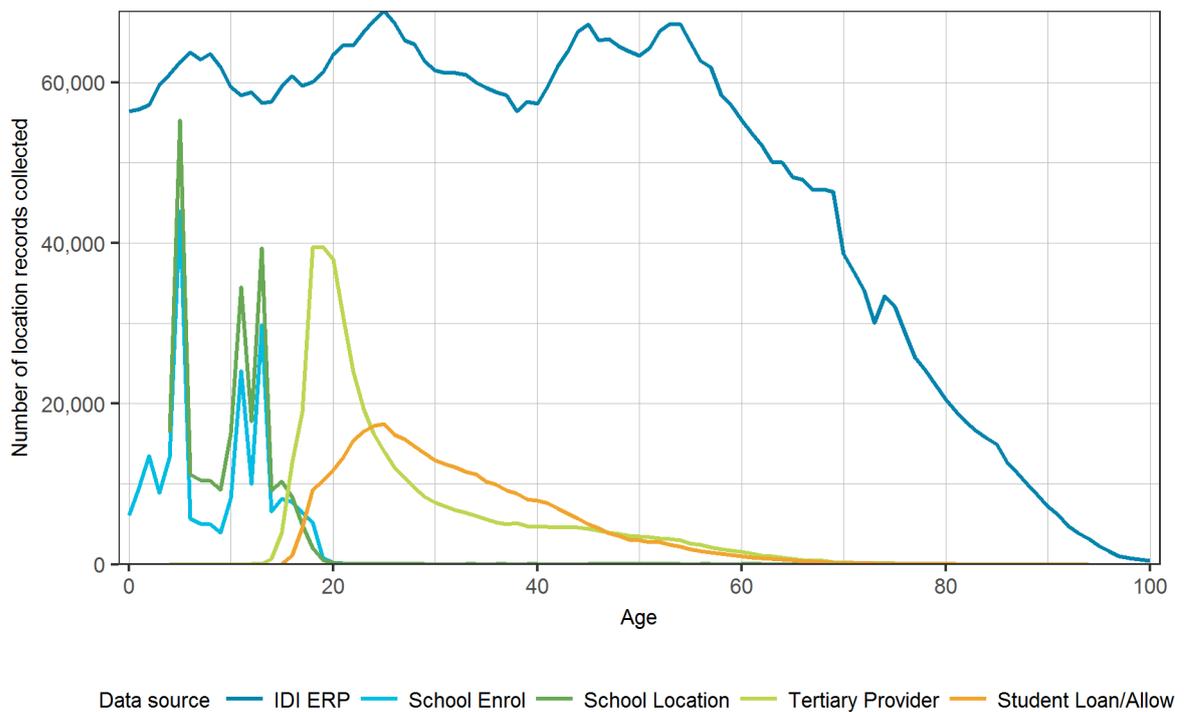


Figure 3: Age at which location records are collected, by record source; employment and welfare sources, and motor vehicle licences, year to June 2016

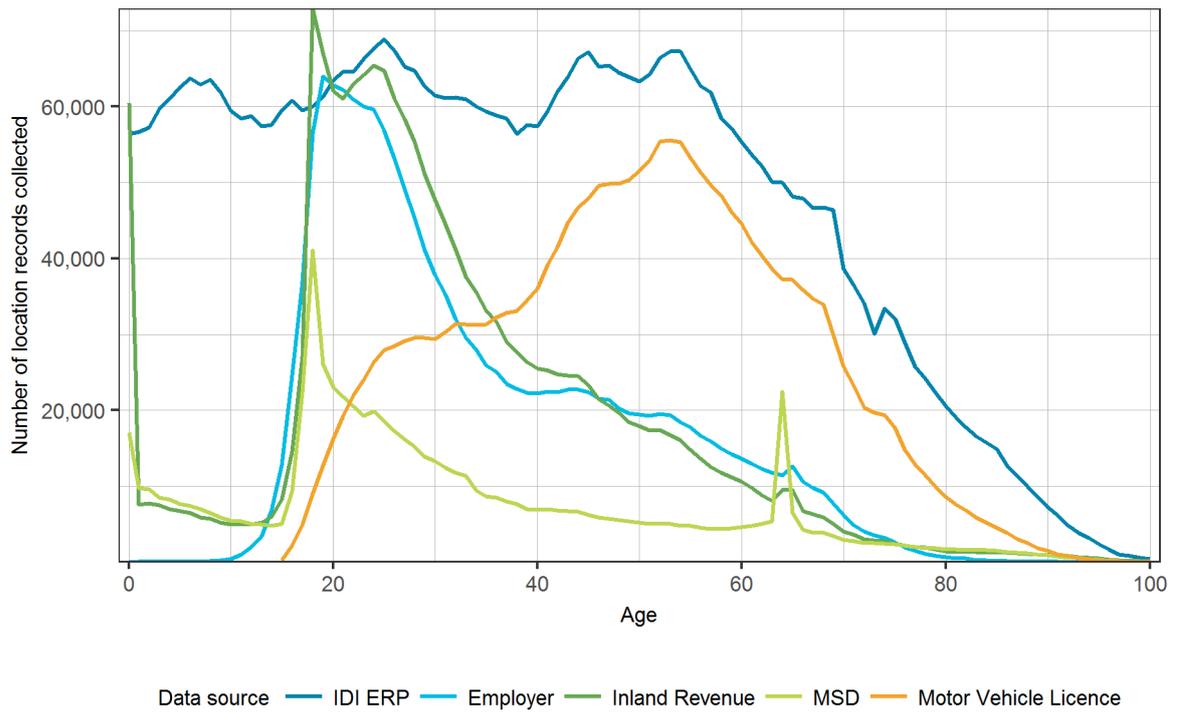
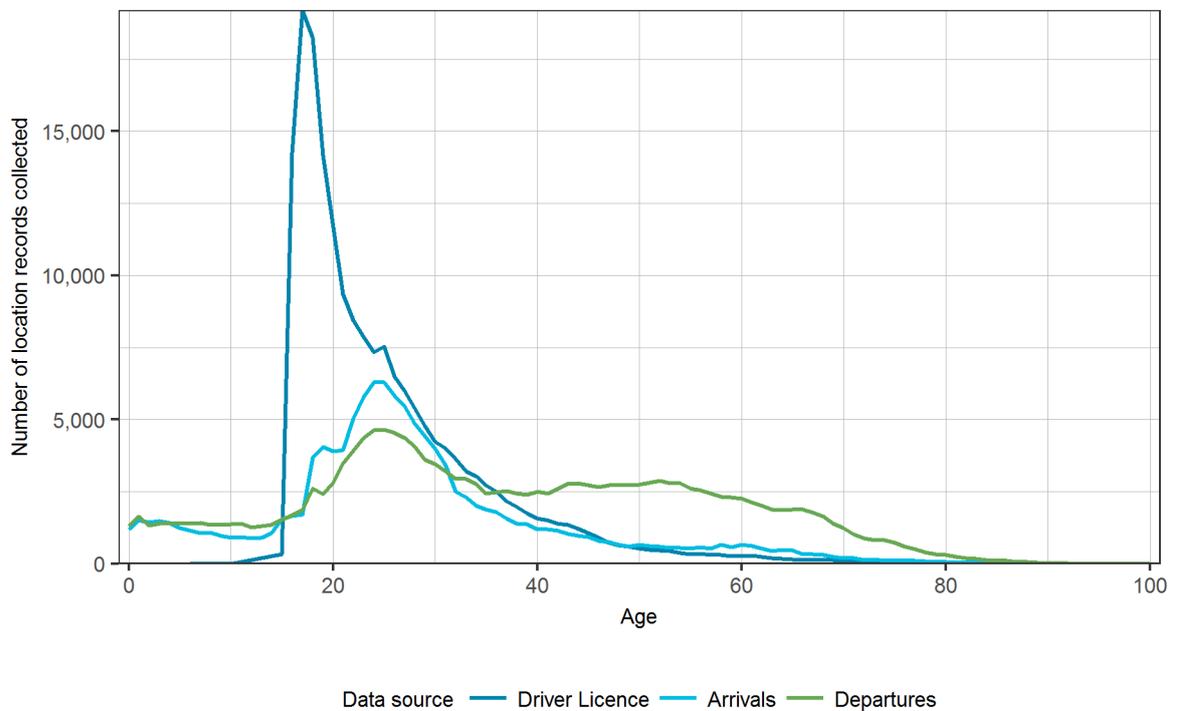


Figure 4: Age at which location records are collected, by record source, other sources, year to June 2016



2.3 Assigning locations to the study population

For most individuals in the population there are multiple potential sources of location information we could select from at any particular point in time. These sources vary however in their inherent quality, in the quality of the match to the IDI, and in the way quality varies across time. Generally speaking the more recent that location information is collected, the more reliable it is, but this is not always the case.

As discussed in section 2.2, while most sources of location data in the IDI are administratively based sources, there are a few survey sources. These sources are likely to provide highly accurate information about people's location at the time they are surveyed, however they are less useful than administrative sources in measuring population change over time.

In the case of Census, measurement is only undertaken in 5-year intervals. While this presents an accurate snapshot, the quality deteriorates across the inter-censal period. Census-based location is least reliable just before a new Census. As a result errors are not consistent across time, and comparisons of different time periods are likely to be biased.

In the case of the HLFS and HES, only a small sample of individuals are selected in any particular period. Given that selection for these surveys is predominantly a random process, they are not likely to provide good scope to fill gaps in coverage of key hard-to-measure population subgroups. In addition, the match rate between these survey sources and the IDI is relatively low.

On the other hand, due to their high accuracy, survey sources present an excellent source of information against which to develop and assess rules for the allocation of individuals to a location.

Allocation strategy

Our allocation strategy uses the 2013 Census to develop a set of selection rules to allocate a location to an individual at any point in time. We then test these rules by assessing how well they match the other survey sources available to us. Testing against the 2008 location information from the 2013 Census allows an assessment of quality at a time when many sources of information were either only recently available (for example school enrolments) or were still likely to be subject to significant under-coverage (for example NHI¹¹). Finally, testing against the HLFS and HES allows us to not only validate the rules against an independent and reliable source of address information, it also allows us to test how robust the rules are to different years of interest.

The strategy we take can be broadly defined as follows:

1. **Extract the 2013 Census population**, and match all potential addresses to that population.
2. **Choose potential locations from each source**, up to two for each person; one being the most recent record prior to the date of interest (5 March 2013) and one the first record subsequent to that date.

¹¹ Approximately double the number of addresses were collected each year from 2008 onwards than were collected from 2004 to 2007 (approximately 2 million instead of 1 million in the earlier period). A particularly large number were collected in 2008 and 2016, with 2.9 million and 2.7 million collected respectively.

3. **Construct allocation rules** to decide which location record is most likely to be correct for any individual based on the potential location records for that person.
4. **Extract an IDI-based resident population and assign locations** to each individual in the population in each period of interest.

Following the construction of the IDI ERP and allocation of addresses, we can then decompose and describe population changes according to the type of change and the characteristics of the population, as described in section 2.4.

Extract the 2013 Census population

As discussed above we assume the recorded 2013 Census location is the ‘true’ residential location of the population in March 2013. We then assess the accuracy of the location rules we develop against this location, and seek to maximise the number of times our estimated location matches it.

The 2013 Census is matched at an individual level to the IDI, with a reasonably high match rate (approximately 94 percent of Census records were successfully matched to the IDI spine¹²) and a reasonably low estimated false positive rate (approximately 1.5 percent of matched records are estimated to be incorrectly matched). This gives us a very large number of records to use to define allocation rules, and a high level of confidence in the addresses used.

For most individuals a large number of location records are held in the IDI, each extracted at a particular date, and generally updated as a result of an administrative event occurring, or as the result of an individual notifying an agency of a change of address.

Choose potential locations from each source

In general we assume that the closer the collection date is to our date of interest, the greater the reliability of that location information. Beyond this however, some sources are inherently more accurate than others, either due to the way they are collected, the way they are coded, or the accuracy with which they are matched to the IDI. For some sources we use the location of a third party (school, tertiary provider or employer) as a proxy for a person’s residential address, a process which introduces significant error in some cases.

For any particular source, we assume that the closer a record collection date is to our date of interest, the more accurate it will be, and we disregard records that are more distant in time. It is unclear however, whether a date collected after our date of interest will be as accurate as an address collected before the date of interest and at the same proximity to that date. For this reason, we consider both the most recent date before the date of interest, and the first one after the date of interest, as candidate locations for consideration¹³.

In total, we have 3.89 million Census records which are linked to the spine and have an address, and 3.76 million where we also have at least one location record from an administrative source. These people have a total of 34.0 million location records

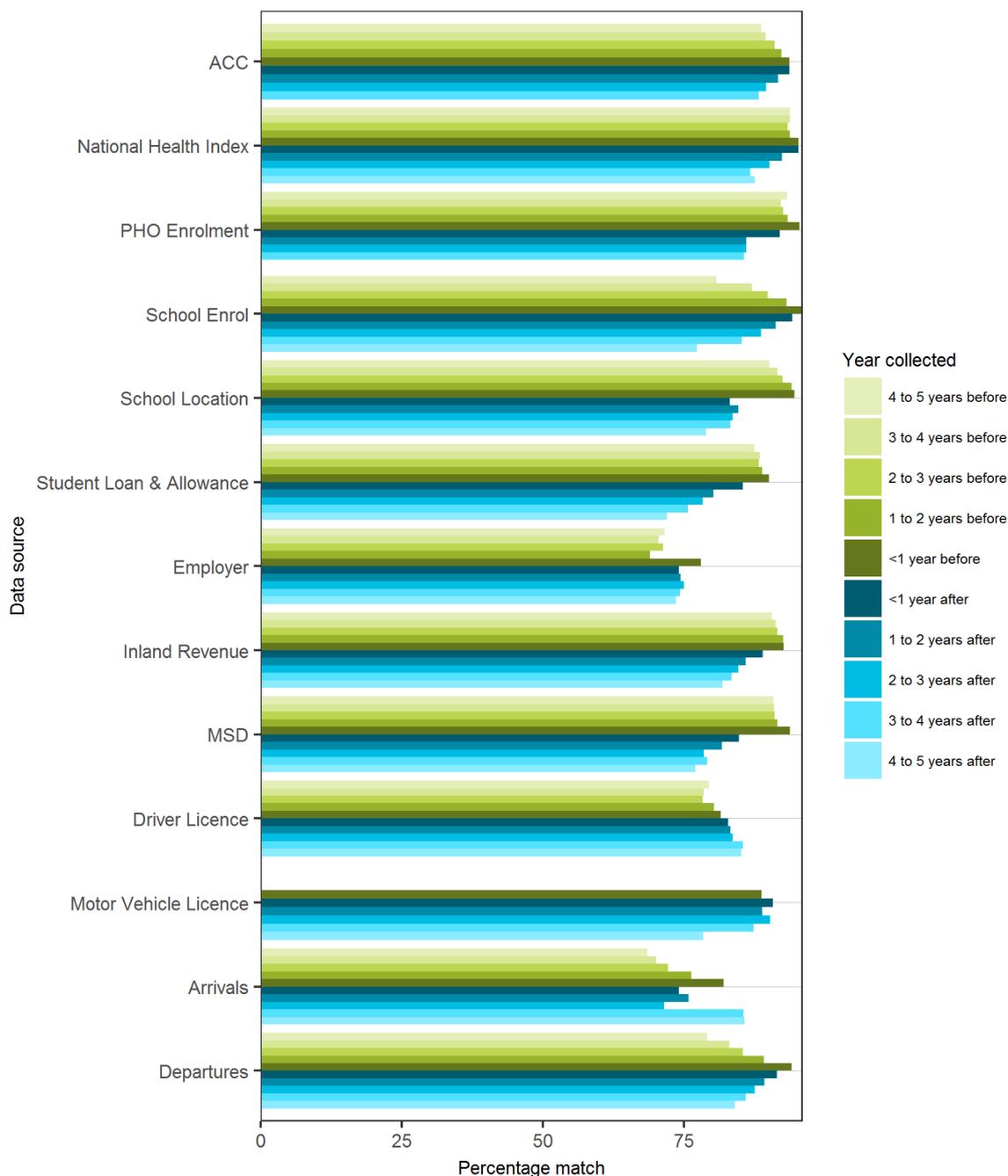
¹² The IDI ‘spine’ is a dataset which all other IDI datasets link to, and which is designed to cover the IDI target population as much as possible. See Black (2016).

¹³ We further exclude any addresses that were collected more than 10 years before, or 5 years after, our date of interest. When looking at the 2013 Census, for most sources we only have information to 2015, 2016, or early 2017, so in practice this latter period is shorter.

associated with them, on average nine records per person, but we need a rule to decide which of these is most likely to be an accurate reflection of each individual’s location on Census night.

The variation in the likelihood of an administrative address matching the ‘true’ Census address is illustrated in Figure 5 below. Almost all address sources have the highest match rate where the lag (or lead) is less than a year before (or after) the date of interest. The one exception to this is drivers’ licensing, where longer lags tend to be more accurate. For drivers licences it may be that the issue date is not the date the address information was most recently updated, and as such addresses for this group may have been updated more recently than the drivers’ licence issue date would indicate.

Figure 5: Accuracy of TA location by source and years since data was collected



Construct allocation rules

We expect proximity to our date of interest to be an important indicator of match quality, and some sources to be inherently higher quality than others. We also know that location information tends to be collected at different ages as indicated in Figure 1 to Figure 4. Where records are collected at other ages, the quality of that location record may be different. As such, we use the source, the number of days between our date of interest and the location record being collected, and the age of the individual to make allocation decisions.

Now that we have a list of potential administratively-sourced locations for each individual at the Census date, and we know whether each of these is correct, we can develop some rules to optimise the choice of location sources that is most likely to correctly match the Census usual residence location. The process consists of two main stages:

1. **Estimate the probability that each potential location record is correct** according to the characteristics of the person, the source of the location record, and the time since the record was collected.
2. **Make a decision on which location to assign**, by either using the record with the highest probability of being correct as estimated in step 1, or the record with the next highest probability.

Estimate the probability that each potential location record is correct

In order to estimate the probability that each location record is correct, independently of other potential location records, we begin by constructing a decision tree. This allows us to assign a probability to each person and potential location. The risk with decision trees is that we 'overfit' our tree to the data. This means that decisions are made based on essentially random characteristics of the data we use to construct the tree, and these decisions cannot be generalised to other samples.

To mitigate this risk we take a tree bagging¹⁴ approach. We construct a large number of trees¹⁵, each from a different bootstrap sample of the population, calculate the resulting probabilities, and then average the probability across all bootstrap replicates. This gives us a fairly stable probability for each potential TA record for each person that indicates the likelihood that TA will match the TA derived from the Census address. The bagged trees are trained on a random 50 percent sample of the population, and are validated on the other 50 percent.

Make a decision on which location to assign

We could just take the record with the highest probability of a match to Census as being the best TA to allocate to an individual, but this approach treats the different potential location records as being independent of each other, whereas in fact they all relate to the same individual. If an individual has one potential location with a high probability, and another location from multiple alternative sources with slightly lower probability, it may be that the location collected from multiple sources is more likely to match the individual's true location. We make the decision on which location to use through a second tree-based

¹⁴ 'Bagging' (short for 'bootstrap aggregating') is an approach designed to minimise overfitting in classification algorithms. It was first proposed by Leo Breiman in the 1990s (see Breiman 1996).

¹⁵ We run our analysis across 100 samples.

decision process. In this process we take the probabilities predicted by the first stage model as inputs to the second stage, in the spirit of 'stacking' (Wolpert 1992).¹⁶

In this second stage we make a decision whether to allocate the TA based on the location record with the highest probability of being correct, or whether to switch to the TA with the next highest probability. This decision is made based on a number of factors which we expect to be linked to an alternative TA potentially being a better match:

- How many potential location records agree with the first TA record?
- How many agree with the first record and have a probability less than 0.05 lower?
- How many agree with the first record and have a probability less than 0.01 lower?
- What is the difference between the probability of the first TA record being correct and the alternative one being correct?
- How many potential location records agree with the alternative TA record?
- How many agree with the alternative record and have a probability less than 0.05 lower?
- How many agree with the alternative record and have a probability less than 0.01 lower?

Using this approach we were able to allocate TA locations to 3.74 million¹⁷ of the 3.89 million Census records which matched to the IDI spine and had an address (96.1 percent). Taking the record with the highest probability from the first stage of TA allocation would have resulted in 94.7 percent of these IDI-based TAs agreeing with the Census TAs (an error rate of 5.3 percent). After applying the second stage, this error rate dropped by around a third to 3.6 percent.

Extract an IDI-based resident population and assign locations

Figure 6 below contrasts the official Stats NZ estimated resident population, with Stats NZ's IDI-based ERP (SNZ IDI ERP), and the IDI-based ERP developed for this study (IDI ERP). Across all years the IDI ERPs are consistently larger than the official ERP, although there has been some convergence in recent years. In 2008 the official SNZ ERP was 4.26 million, compared to an SNZ IDI ERP of 4.33 million and an IDI ERP of 4.30 million. More recently, the three ERPs have been closer, with an SNZ ERP of 4.69 million, and IDI ERPs of 4.71 and 4.70 million.

Differences between the two IDI ERPs are likely be largely due to different definitions of usual residence. These are likely to influence the populations in competing ways. Our residence definition is based on time in New Zealand over a fixed period, which is likely to make the estimate lower, while our longer activity period is likely to make it higher. Similarly, differences between the SNZ ERP and the IDI ERPs are likely to at least partly result from differences in the application of the 'usual residence' definition in the sources which contribute to the estimates. The SNZ ERP may also under-count some groups in the population, despite attempts to account for Census non-response and changes in the population over time. Finally, the IDI may over-estimate the population if individuals are not successfully matched and are treated as being more than one person in the data.

¹⁶ For the second stage a single decision tree is formed. This is trained using half of the validation sample from the first phase to train and the other half for validation (two 25 percent samples from the original population).

¹⁷ This is slightly smaller than the 3.76 million individuals who were reported earlier as having at least one location record in the IDI data, as some of these were from lower quality sources which we excluded from consideration.

Figure 6: Comparison of different estimated resident populations over time

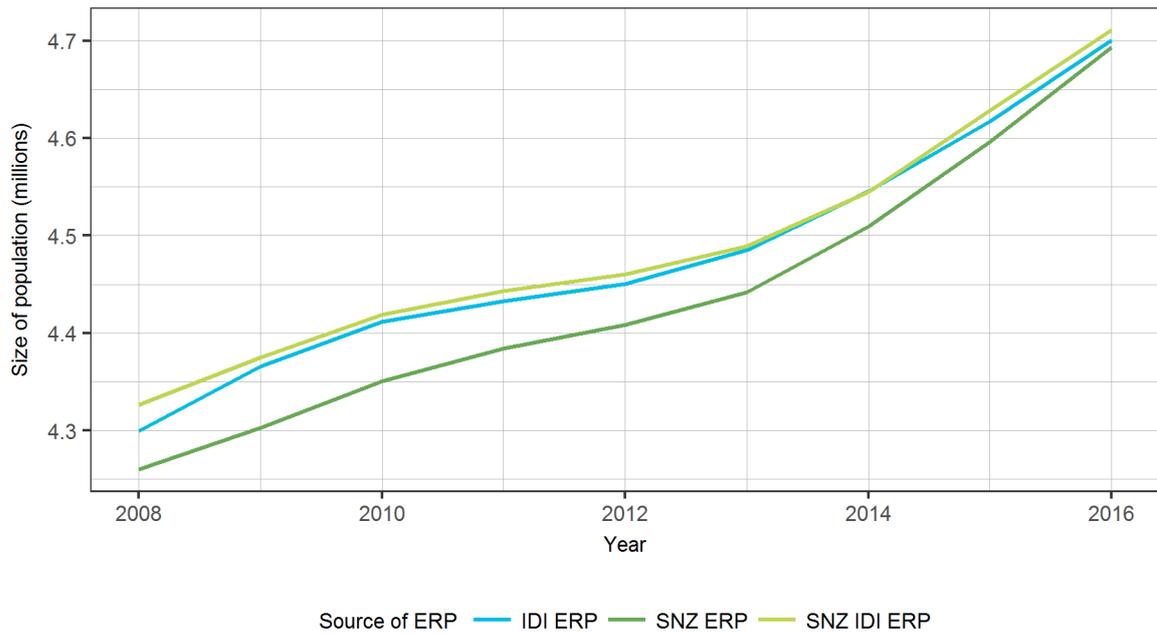
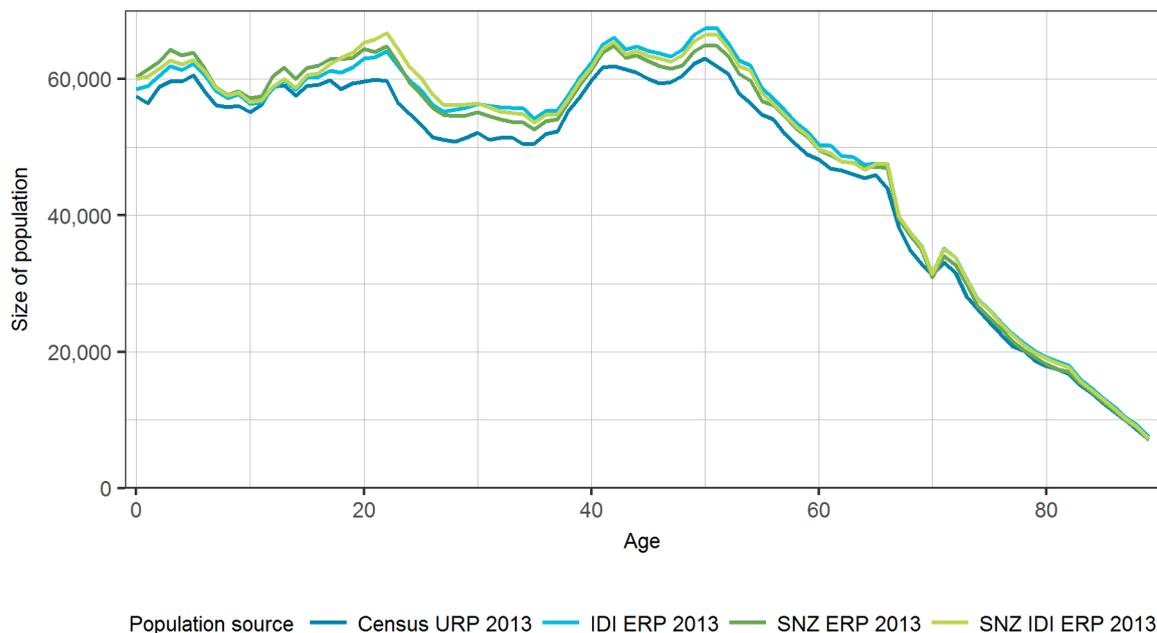


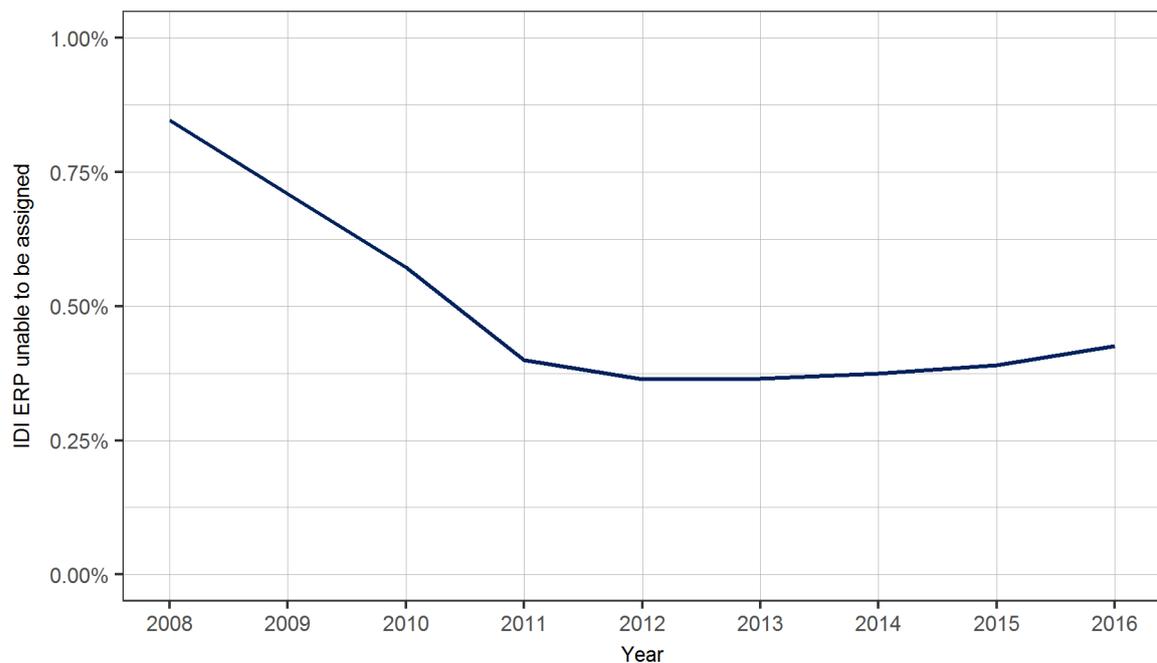
Figure 7 shows a comparison of ERPs and the Census usual residence population (URP) by single year of age. As expected the Census URP is smaller than the estimated resident populations, which are similar to each other. Generally speaking the SNZ ERP is larger than the IDI ERPs at childhood ages, and slightly lower from the mid-20s to the mid-50s. The two IDI ERPs closely follow each other at most ages, however the SNZ IDI ERP is somewhat higher than our IDI ERP through the late teens and early-to-mid twenties. Over these ages our more restrictive residence definition is likely to have a greater impact, as this more mobile population is more likely to spend time overseas, and be excluded on that basis.

Figure 7: Comparison of different 2013 resident populations by single year of age to age 90



Once we have defined our IDI ERP, we wish to assign each individual in the population to a location at June of each calendar year of interest. We are able to do this for almost everyone in the IDI ERP population. Figure 8 shows the percentage of the population we are unable to assign a location to. In every year this is less than one percent, with rates of less than half a percent from 2011 onwards. The rate is higher in the earlier period due to limitations in some administrative data in that period, and slightly higher in the most recent year due to the limited availability of records covering the period after the date of interest.

Figure 8: Percentage of IDI ERP unable to be allocated to a TA, 2008 to 2016



2.4 Decomposing and describing population change

Population change decomposition

Changes in the population are defined as being due to one of three causes: internal migration (also described as migration within NZ), external migration (also described as international migration), and natural increase (births minus deaths).¹⁸ We compare any combination of years from 2008 to 2016. Where a person is in the population of a TA in one year, but not in the population of the TA in the other year, that movement can be described as being an inflow or an outflow, and assigned to one of these three categories, depending on whether the person moved from one TA to another, whether they were born or died between the two periods, or whether they were resident outside of New Zealand (ie they did not meet the 12/16 residence rule in either of the periods).

International migration

Although similar, the 12/16 rule is applied in a different way in this study than in the official observed migration series developed by Stats NZ in recent years. In the official series, to be considered as a migrant arrival a non-resident person needs to spend at least 12 months out of the 16 following months in New Zealand. To be considered as a migrant departure a New Zealand resident must spend at least 12 out of the following 16 months away from NZ.

This means that once someone is considered to be a resident by this rule, they will maintain this status as long as they, at each departure from NZ, spend more than 4 out of the following 16 months in NZ. Similarly, once NZ residents have departed and spent at least 12 months out of the following 16 months away from NZ (i.e. classed as migrant departures), then when returning to NZ, they must spend at least 12 out of the following 16 months in NZ before being considered as a migrant arrival.

Consider Person A and Person B below. Each coloured square symbolises a 2-month period spent in New Zealand, while each blank square represents 2 months spent overseas. Both people are considered to be resident through to December 2010. At that point both depart, but as person B spends fewer than 12 of the next 16 months away from NZ, their resident status is unchanged and they are considered to have not left NZ.

Figure 9: Residence classification example, 12/16 rule



Even though both people were resident in 2010, the fact that Person A was away for 12 months in the 16 months following their December 2010 departure means they are considered to have departed according to the official 12/16 rule. This then conditions their subsequent treatment, so, despite spending the same time as Person B in NZ in 2012, 2013 and 2014, Person A is considered to still be non-resident in those years, while

¹⁸ Where subgroups of the population are defined according to characteristics that can change over time (e.g. age, employment and ethnicity), populations may also change as these characteristics change. The only time such characteristics are explored in this study is with respect to age. The Insights tool includes population pyramids defined in 5 year age groups, and ageing is added as an additional component of change (combined with natural increase). We do not have a way of robustly tracking people's changing ethnic identification, so ethnicity is treated as a fixed characteristic for the purpose of this study.

Person B is considered to be still resident in New Zealand as they are considered to have never left.

Under our implementation of the 12/16 rule, both person A and person B are considered to be in the IDI ERP in 2010, but not in the subsequent three years, as they do not spend sufficient time in the country in the 16-month period centred on each calendar year. They are recorded as a migrant departure from 2010 to 2011, and are not considered to have returned to the NZ IDI ERP by 2014.

As we are interested in using the data for this study to understand the impacts of different components of population change on the local economy and society, it is helpful that we have a consistent definition of residence that is not conditioned on historical travel patterns (as the official 12/16 rule is). In the example above, we might expect person A and B to have the same impact (all other things being equal) on the area in which they live from 2012 to 2014, as they spend the same time in the country over that period, and it is useful that they are treated in the same way in those years.

Another difference between the approaches is that the 12/16 rule follows individuals over a period following and preceding an arrival or departure. We instead focus on a fixed time frame centred on our date of interest. The results presented in this report, and in the Insights tool, use a calendar year. People are considered resident during a particular year if they spent 4 months or fewer absent from New Zealand during the 16 months period incorporating the calendar year and the two months either side of it.

Migration within NZ

Where a person is identified as being located in one TA in one period and another TA in another period they are considered to be internal migrants. Accurate measurement is dependent on timely recording of changes of address in our key data sources. The main characteristic for internal migration is the TA (ie when people leave a particular TA, where do they tend to move to, and how many people move in the opposite direction?). At a total New Zealand level this change nets out to zero, and there is no internal migration component as a result.

Natural increase

Natural increase is defined as the number of births into the population less the number of deaths. Our measure is similar to the official statistical measure, but not all people who were born or died in New Zealand are represented as births or deaths in our data. This is because of the IDI ERP 12/16 rule. If a person spends more than 4 months outside of New Zealand then they are not included in the population for that year, and are excluded from the population decomposition.

For births, this means that children who are born in NZ, but subsequently spend more than four months out of the country, are considered to not have entered the population in that year. If they spend 12 months or more of the following calendar year (in fact the 16 month period from November of the previous year to February of the next), they will be considered a migrant and be included in the international migration figures.

For deaths, where a person spent more than 4 months outside New Zealand in the year they died, they are considered as being a migrant departure in that year, and are not captured as a death in any year.

Residual

There are a small number of people (less than a percent in any particular year, for whom no TA is able to be assigned. This effectively represents a residual in our population decomposition. We may be able to identify at a Total NZ level whether they arrived from or departed overseas, were born or died, but we do not know where they are located in the country, nor whether they moved location. In some cases we are able to assign a person to a TA in one period, but not in the other (generally earlier) period, meaning that we can assign them to an area, but are unable to determine whether they arrived from another TA or were already in the TA in the earlier period. We assign these flows to a residual category. Given the small number of people who are unable to be assigned to a TA, these residuals are generally small, but are somewhat larger in earlier years than later years.

Further describing population change

The IDI contains a wide range of data collected from numerous sources. The production of the data described in this paper opens up the possibility of describing the changing characteristics of the New Zealand population, and the drivers of this change, in many different ways. For the purpose of this study we focus on a few key demographic characteristics, as well as some broad information about the visa category and country of origin of migrant groups.

The characteristics included in the data and reported through the Insights tool are:

- Sex
- Age¹⁹
- Ethnic group
- Territorial authority (TA)
- Visa type
- Country of origin.

Ethnic group

Ethnic group is based on the prioritised ethnicity classification developed by Stats NZ for use in the IDI, and is derived from multiple data sources. A person can be recorded as having multiple ethnic groups and these are reported individually. As such, reported numbers are likely to add to more than 100 percent. Broad categories are European, Maori, Pacific Peoples, Asian, MELAA (Middle Eastern, Latin American, and African) and Other.

Territorial authority (TA)

All measures are presented at both a Total NZ level and for each TA individually. TA is also used to describe migration within NZ. For each TA, movements are identified and quantified to and from every other TA. Chatham Islands Territory is excluded from all analyses however, as people are unable to be consistently allocated to the Islands using IDI. IDI-based population figures for the Chatham Islands are much smaller than official statistics population figures.

¹⁹ We generally report age in 5 year bands.

Visa type

Broad visa types reported are Work, Student, Resident and Other. This is based on the most recent visa approved for a person in the last year over which changes are being examined (e.g. if we are looking at population change between 2012 and 2016, the category will be based on the visa most recently issued prior to the end of December 2016). The 'Other' category is largely made up of people travelling on an Australian passport, who have a right to reside in New Zealand without a work, student, or residence visa, but also includes some people who are in New Zealand long-term on a Visitor's or other temporary visa. Finally, a fifth visa category of "NA (NZ country of origin)" includes travellers classified as having a New Zealand country of origin (as discussed below).

Country of origin

Classification by country is complicated by the large number of definitions that could potentially be used (e.g. country of birth, country of citizenship, country of residence) and by lack of data to apply some of these definitions. For example, country of birth is not collected for the complete resident population in the IDI, and may not be a relevant classification for people who may have been born in another country but lived most of their life in New Zealand.

Citizenship is also difficult to use, as citizenship data is currently not available in the IDI. Data from passports people use when they arrive in or depart from New Zealand is available, and this can be used to infer a country of citizenship, but this is not comprehensive. We take the country of origin in our classification from the passport used in a person's first observed arrival in New Zealand.²⁰ We then re-classify a person as having New Zealand country of origin if they were granted residence more than 10 years before June of the last year over which changes are being examined, or where they first arrived on an Australian passport in the same period. Countries of origin were grouped within regions, with the main countries within each region being classified and reported separately. Table 1 shows how countries were classified.

²⁰ Arrivals and departures are observed from 1997, meaning that we are able to observe at least a ten year travel history for all individuals during our period of interest (from 2008 to 2016).

Table 1: Country and region of origin classification

Region	Country
New Zealand Oceania	New Zealand Australia Fiji Samoa Tonga Other Oceania
North-West Europe	France Germany Ireland Netherlands United Kingdom Other North-West Europe
Southern and Eastern Europe	Russia Other Southern and Eastern Europe
North Africa and the Middle East	North Africa and the Middle East
South-East and North-East Asia	China Hong Kong Japan Republic of Korea Malaysia Philippines Taiwan Thailand Viet Nam Other South-East and North-East Asia
Southern and Central Asia	India Sri Lanka Other Southern and Central Asia
The Americas	Brazil Canada United States of America Other Americas
Sub-Saharan Africa	South Africa Other Sub-Saharan Africa
Other	Other

2.5 Quality assessment

Individual assessment against 2013 Census

The main sources of information about the quality of our IDI-based assignment of the population to TAs are official survey sources, particularly the Census. The Census is almost a full enumeration of the New Zealand population, and is one of the few sources of survey data in New Zealand that follows the movement of people over time, albeit over specific five-year periods. This enables us to check the quality of the IDI location assignment not just at a single point in time, but also across two time points.

Other survey sources of data also linked into the IDI are the Household Labour Force Survey (HLFS) and the Household Economic Survey (HES). These provide a source of data to validate our results against. As these are ongoing data collections, they also allow us to better understand the stability of the allocation to TAs over time. This is particularly important given that we define the allocation rules using a population defined at a single time point (the Census date of 5 March 2013).

Address quality by source and time of collection

Table 2 shows the number of location records that were assigned using different sources of address data, and how well each of these aligned with the Census location. In assigning TA locations to individuals responding to the Census, health sources were by far the most frequently used, with the NHI being the source of over half of locations assigned (52 percent) and PHO enrolments the source of over a quarter. These were also two of the highest quality sources, at 97 percent. The only higher quality sources were school enrolments (97-98 percent), although as expected these only accounted for a relatively small percentage (around 4 percent) of all locations assigned.

Table 2: Use and quality of location records by data source

Location record source	Number of records	% of locations assigned	% match with Census TA
School Enrolment	89,139	2.43	98.2
School Location	54,240	1.47	97.4
National Health Index	1,917,285	51.74	97.3
PHO Enrolment	1,041,522	27.98	96.8
ACC	125,931	3.31	94.7
Tertiary Provider	94,743	2.47	94.0
MSD	48,003	1.25	93.5
Inland Revenue	340,734	8.68	91.8
Student Loan/Allowances	9,135	0.23	90.4
Departures	5,292	0.13	88.2
Motor Vehicle Licence	7,644	0.19	88.1
Arrivals	2,808	0.06	77.9
Driver Licence	2,205	0.05	76.3
Employer	1,812	0.03	68.9
Total	3,740,493	100.00	96.4

In general the higher quality sources accounted for greater shares of the locations assigned, as to be expected given our prioritised allocation approach. The next four sources in terms of quality, were ACC, Tertiary provider location, MSD, and Inland

Revenue, making up a combined 16 percent of all addresses assigned and each with a 92 percent or higher match with the Census location. The 6 lowest quality address sources only accounted for 0.7 percent of locations assigned, although this still represented almost 30,000 people who otherwise would not have been allocated a location.

Table 3 presents similar information to that presented above, but broken down by the time since the location record was updated, instead of the source. Over 80 percent of locations assigned were derived from information collected less than a year after the date of the Census, and up to 3 years before the Census. Another 15 percent were collected 3 to 6 years before the Census, with the remaining 3 percent being collected more than a year after the Census or more than 6 years before it. Many sources of data were not available during these more distant periods, but even where they were available, the quality was much lower. In general addresses with the closest proximity to the Census date, especially those collected before the Census, were of higher quality. Around 96 to 97 percent of TAs recorded in location records collected in the 6 years leading up to the Census matched the Census TA.

Table 3: Use and quality of location records by time since record collected

Location record collected	Number of records	% of locations assigned	% match with Census TA
4 to 5 years after	1,167	0.03	68.6
3 to 4 years after	7,326	0.20	80.2
2 to 3 years after	12,036	0.32	83.4
1 to 2 years after	10,146	0.27	83.7
Less than 1 year after	403,314	10.78	94.2
Less than 1 year before*	1,648,014	44.06	96.8
1 to 2 years before	660,786	17.67	97.0
2 to 3 years before	358,401	9.58	97.3
3 to 4 years before	293,685	7.85	96.7
4 to 5 years before	191,532	5.12	96.5
5 to 6 years before	61,239	1.64	95.9
6 to 7 years before	26,184	0.70	94.5
7 to 8 years before	17,493	0.47	92.7
8 to 9 years before	20,292	0.54	92.7
9 to 10 years before	22,641	0.61	93.4
10+ years before	6,228	0.17	88.0
Total	3,740,493	100.00	96.4

* Also includes location records collected on the day of interest.

Address quality by demographic characteristics and geographic location

There was some variation in location matching across different TAs for Census respondents, but for all TAs over 90 percent of records were able to be allocated a location in IDI; and in all but one TA, over 90 percent of records were allocated the same TA in both Census and IDI. Where locations were unable to be assigned, this was primarily because the Census record was not successfully linked to the IDI spine.

The highest likelihood that a location record could be assigned from administrative sources was for Census respondents in Upper Hutt City (97.4%), while the lowest was in

Ōpōtiki District (92.3%).²¹ Of the location records that were assigned, the highest accuracy was in Auckland (98.7% of Census records having the same TA allocated from IDI) while the lowest accuracy was in Mackenzie District (87.4%). These are the largest and smallest TAs according to population, respectively, however size was not always predictive of better quality location matching. While the main centres generally had match rates in excess of 95 percent, many smaller areas (such as South Waikato, Whakatane, South Taranaki, and Grey Districts) had similarly high rates.²²

Young adults tend to be highly mobile, and as such are known to be difficult to locate and to collect data from. As illustrated earlier we attempt to use a number of data sources that collect data from young people in their late teens and early twenties. In particular, large amounts of location data from ACC claims, tertiary providers, student loans and Inland Revenue (tax, and student loans and allowances), employment, MSD and driver licensing, are collected from people in this age range, complementing high quality NHI and PHO sources.

The number and quality of matches are broken down by both age and sex in Table 4, as the patterns are somewhat different for males and females, although in general females were both more likely to have a location assigned to them, and more likely to have the address matched the Census address.

Table 4: Location assignment and quality by sex and age

Sex and age group	% with locations assigned	% match with Census TA
Male		
Under 15	95.9	97.1
15 to 24	95.0	92.7
25 to 44	95.4	94.9
45 to 64	95.8	97.0
65 and over	95.7	98.3
Total male	95.6	96.1
Female		
Under 15	96.3	97.0
15 to 24	96.9	93.1
25 to 44	97.3	96.3
45 to 64	96.8	97.6
65 and over	95.0	98.3
Total female	96.6	96.7

For males, as we expect, young adults are most difficult to locate, although we are still able to assign a location to 95.0 percent of 15 to 24 year olds (compared to 95.6 percent for males overall). The quality of the matching is much lower than for other age groups, consistent with a more mobile population: 92.7 percent were successfully matched to the Census location, compared to 96.1 percent for males overall.

Although young women aged 15 to 24 were more likely to be assigned a location (96.9 percent) than young men, the location was only a little more accurate than for men.

²¹ Chatham Islands Territory had lower location assignment of 85.2% and had extremely low accuracy of 39.7%. It was excluded from our analysis.

²² This is consistent with an expectation that the estimated accuracy for small populations is likely to have more random variation than for large populations.

Older women, aged 65 and over, were also relatively difficult to assign a location to, with only 95.0 percent having a location assigned, although the accuracy of the match was extremely high for both men and women in this age range.

Validation against other sources

Reported 2008 location from 2013 Census

In the 2013 Census, people were asked for their usual residence address five years earlier. This allowed Stats NZ to produce estimates of internal migration over the five year period, although these are limited according to the quality of the responses. Many people either did not provide a useable response, or provided a response that could only be coded at TA or region level. Nevertheless, these responses provide a useful validation of the allocation rules we developed using the 2013 Census location.

We are interested in the degree to which people's Census-reported 2008 TA matches our IDI-based TA allocation. Table 5 shows how many 2008 location records were able to be assigned a TA from IDI, and the degree to which the IDI-based TA matched that reported in the Census. Overall, of the almost 3.3 million Census records which could be assigned to a 2008 TA, 95.9 percent could also be assigned an IDI-based TA, only a little lower than the 96.1 percent for the 2013 location.

Table 5: Location assignment and quality of 2008 location assignment

2013 Census location 5 years ago (March 2008)	Number of records	% of locations assigned	% match with Census TA
Same TA as 2013	1,762,134	96.6	96.8
Different TA from 2013	1,511,550	95.2	89.2
Detailed 5 years ago address collected	3,207,264	96.0	93.7
Only 5 years ago TA collected	66,426	92.6	76.5
Total	3,273,684	95.9	93.3

The table also splits the 2008 location records according to whether the person moved TAs between 2008 and 2013. People who are mobile are likely to be more difficult to locate. Nevertheless, locations could be assigned to 95.2 percent of records for those who moved TA (almost half of all people who were assigned a TA at both times), and the quality was still reasonably high, at 89.2 percent.

Finally, we separated out those respondents (around 66,000) who could be assigned to a TA, but not to a more detailed geographic location. Of that group, only around three-quarters were allocated to the same TA in the IDI, possibly indicating that the quality of the Census-based TA location was not as good for that group.

Reported location from survey sources

We were also able to validate our TA allocation rules against two high-quality survey sources, the Household Labour Force Survey (HLFS) and Household Economic Survey (HES). Although these sources of data are of high quality, the match to the IDI is less reliable. Many respondents were not able to be linked to the IDI spine, and as a result only around 80 percent of HES respondents and two-thirds of HLFS respondents could be assigned to a TA based on the IDI (see Table 6). The percentage of respondents who could be assigned a TA was largely driven by the percentage of records that are successfully matched to the IDI spine. The match rate for HLFS is only around 70 percent.

The quality of allocation to TAs, where this was able to be done, was high, with over 96 percent of IDI-based TAs matching HES TAs in every year apart from 2006 and 2008. The HLFS match was lower, at around 93 percent, however this is likely to be due to poor matching to the IDI rather than any issues with the IDI-based allocation of TAs. A very positive result is that the quality of TA allocation based on rules developed from the 2013 Census seems to be extremely stable across time (particularly since 2008), indicating that we can confidently use these rules to allocate TAs to the IDI ERP over our period of interest, from 2008 to 2016.

Table 6: Location assignment and quality of HLFS/HES respondents over time

Survey Year	Household Economic Survey		Household Labour Force Survey	
	% of locations assigned	% match with survey TA	% of locations assigned	% match with survey TA
2006	82.8	95.2	38.5	88.7
2007	79.7	96.3	60.5	90.2
2008	80.4	95.8	64.7	92.7
2009	82.8	96.5	64.8	93.1
2010	86.5	96.7	66.2	93.4
2011	84.3	96.8	65.7	93.1
2012	84.9	96.6	65.4	93.1
2013	83.6	96.9	64.2	93.5
2014	80.3	96.4	62.8	93.9
2015	79.1	96.5	62.8	92.9
2016	n/a	n/a	66.2	94.8

Movements between 2008 and 2013

While our allocation results in 96.4 percent of IDI-based TAs as at March 2013 matching the TA recorded in the 2013 Census, the ultimate aim of this work is to decompose changes in the population. This means that we not only want to accurately locate the majority of people at a point in time, we also want to accurately record their movements within New Zealand, so we can estimate rates of internal migration between TAs.

In order to assess the quality of the TA allocation across multiple time periods we look at the approximately 3.0 million people who were assigned a TA in both 2008 and 2013 from both Census and IDI. We then look at the degree to which they are identified as internal migrants in both sources, the degree to which they are allocated to the same two TAs in 2008 and 2013, and the degree to which people who are mobile or not mobile are accurately able to be matched in 2013.

Table 7 shows the degree to which people who moved TAs according to the Census also moved TA according to the IDI. This shows that around 300,000, or three-quarters of all Census movers were also identified as IDI movers. Interestingly both sources identified that a total of around 400,000 people in this data moved over the two periods.

Table 7: Matching of Census movers and IDI movers

	IDI non-mover	IDI mover	Total
Census non-mover	2,532,732	97,611	2,630,343
Census mover	101,580	306,621	408,201
Total	2,634,312	404,232	3,038,544

Table 8 shows the degree to which Census non-movers are successfully allocated to the correct TA in the IDI, and the degree to which movers are allocated to the correct pair of TAs, according to the Census. While over 90 percent of all people in our data were successfully assigned to the same two TAs in both the Census and IDI, there was (not surprisingly) a far better match for non-mobile respondents (96 percent, compared to 63 percent).

Table 8: Matching of 2008 and 2013 TA for Census movers and non-movers

	Non-match	Match	% match
Census non-mover	114,915	2,515,428	95.6
Census mover	150,141	258,060	63.2
Total	265,056	2,773,488	91.3

While the IDI seems to identify a similar internal migration rate to Census, it doesn't seem to be able to accurately allocate mobile people to the correct TAs in both periods. This could be due to either time lags or inaccuracies in the IDI source datasets. Care should be taken in undertaking research which is reliant on identifying internal migration at an individual level using IDI. While movements between areas may be reliable on aggregate, the IDI does not seem to provide accurate data about the timing of an individual's movements. Such movements may be identified subsequent to their actual occurrence, may be missed, or may even be inferred as occurring in the incorrect direction.

Finally, Table 9 looks at the degree to which mobile people are able to correctly be assigned to a 2013 TA. As above, movers are more difficult to allocate to a TA, with only 85 percent of movers being successfully assigned to the correct 2013 TA, compared to 98 percent of non-movers.

Table 9: Match to 2013 location for Census movers and non-movers

Census move by whether IDI had the same TA in 2013

	Non-Match 2013 TA	Match 2013 TA	% match
Census non-mover	41,280	2,589,066	98.4
Census mover	63,156	345,042	84.5
Total	104,436	2,934,108	96.6

Aggregate comparison against official sources

As indicated above the IDI-based TA allocation seems to be high quality, albeit of higher quality for people who are less mobile. Nevertheless, the most important thing for our purpose is that aggregated population change statistics are accurate. In this section we compared aggregate population change statistics from the IDI with official estimates.

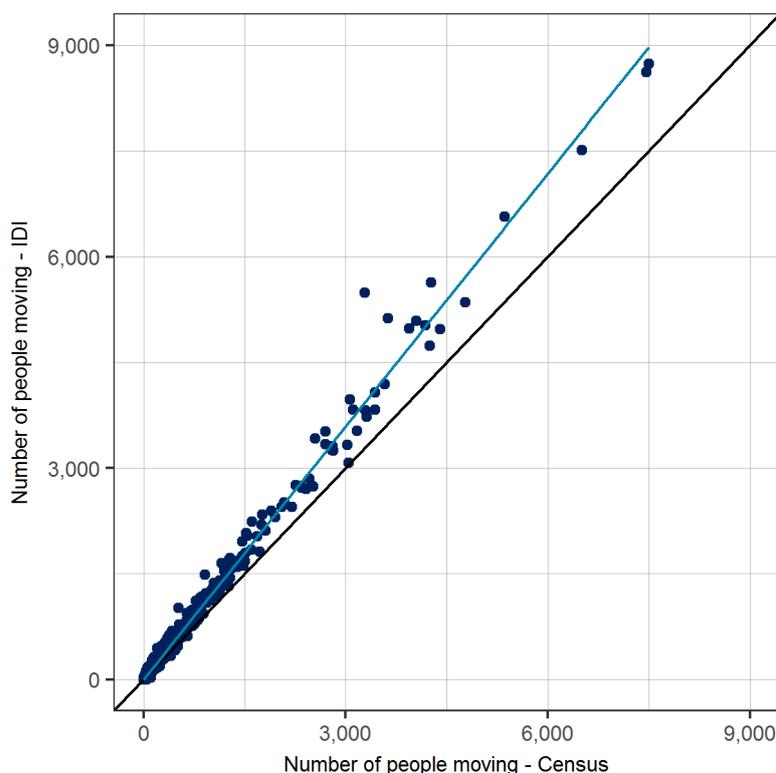
Migration within NZ

Official statistical measures of internal migration between New Zealand TAs are calculated using Census data, with people asked to report their usual residence five years previous to the Census. Of the 3.27 million people who were assigned a TA in 2013 and five years previously in the 2013 Census, 457,000 moved between TAs over the period. With our data, we identify a larger population who were resident in NZ TAs in both periods – a total of 3.80 million people. The lower Census figure is not unusual as the Census figures do not include people who did not respond to the Census, or who did not provide a

useable 5-years-ago address. Although we identify a larger population, the internal migration rate²³ is similar between both sources, with an estimated 14.5 percent migration rate according to IDI, and a 14.0 percent rate according to Census.

Figure 10 shows internal migration flows between 2008 and 2013 as measured in the 2013 Census and using the IDI. Each point represents a pair of TAs, with the x axis showing the number of people moving from the first TA to the second TA according to the Census, and the y axis showing the number of people moving in the same direction and between the same TAs according to the IDI. For each combination of TAs there are two points on the graph – one showing the number of people moving from the first area to the second, and the other showing the number of people moving in the other direction.²⁴

Figure 10: Internal migration flow comparison of Census and IDI-based figures – All combinations of TAs 2008 to 2013



A point would lie on the 45 degree line if the Census and IDI both showed the same number of people moving between the two TAs represented by that point. As expected given the net Census undercount, particularly for 2008 address information, IDI-based measures of migration within NZ are generally higher than Census-based measures, and the points generally lie above the 45 degree line as a result. Nevertheless, there is a strong linear pattern evident in the data, reflected in the close clustering of points around the regression line on the plot.²⁵ Only a few points diverge markedly from the regression line, the most noticeable being that relating to the number of people moving from

²³ In this case we calculate the internal migration rate as the number of people moving into a TA over a period of interest divided by the population who were identified as being in New Zealand in both periods and were able to be allocated to a TA in both periods. This is different from the internal migration rate presented using IDI data in other parts of this report, and is only being used for comparisons with Census figures.

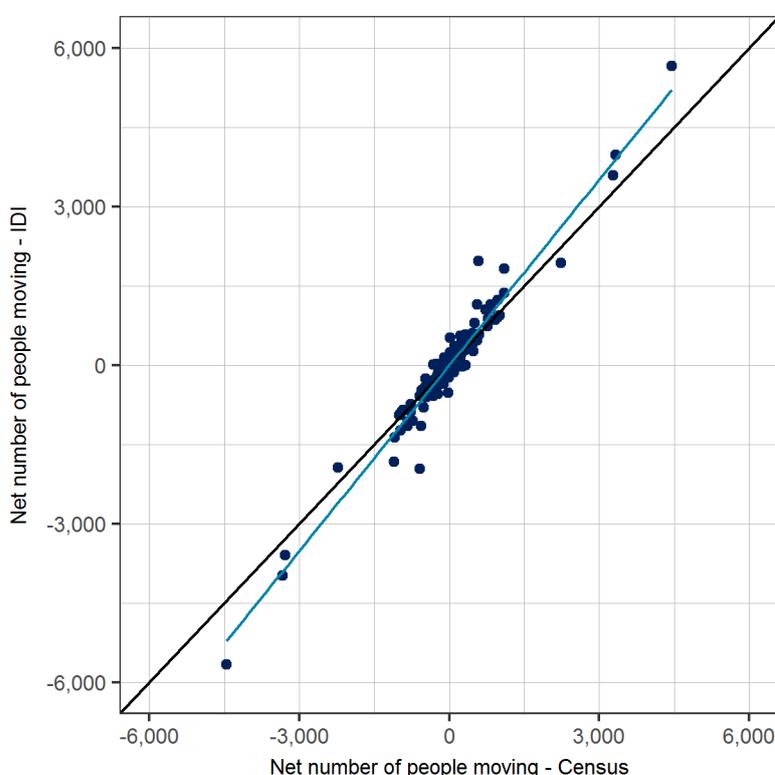
²⁴ In total the figure has 4,290 points, representing all combinations of the 66 TAs represented in the data.

²⁵ This is confirmed by a Pearson's correlation coefficient of 0.996, very close to 1.

Far North District to Auckland; a number estimated at 3,282 according to Census, but 5,487 according to the IDI.

Figure 11 shows similar information, but for net flows, instead of gross flows. This shows the number of people moving between each pair of TAs, minus the number of people moving in the other direction. This means that figures can be positive or negative. By construction the lower left-hand quadrant of the plot is a mirror image of the upper right-hand quadrant – each positive net flow is matched by an equivalent negative net flow in the other direction. Again the 45 degree line provides an indication of the similarity of the Census-based and IDI-based measures. As with the plot of gross flows, there is a strong relationship between both measures of net flows.²⁶ Again the regression line is steeper than the 45 degree line, and flows between the Far North and Auckland show the greatest divergence between the two measures; a net flow of 1,965 people from the Far North to Auckland according to the IDI, but only 585 according to the Census.

Figure 11: Net internal migration flow comparison of Census and IDI-based figures – All combinations of TAs 2008 to 2013

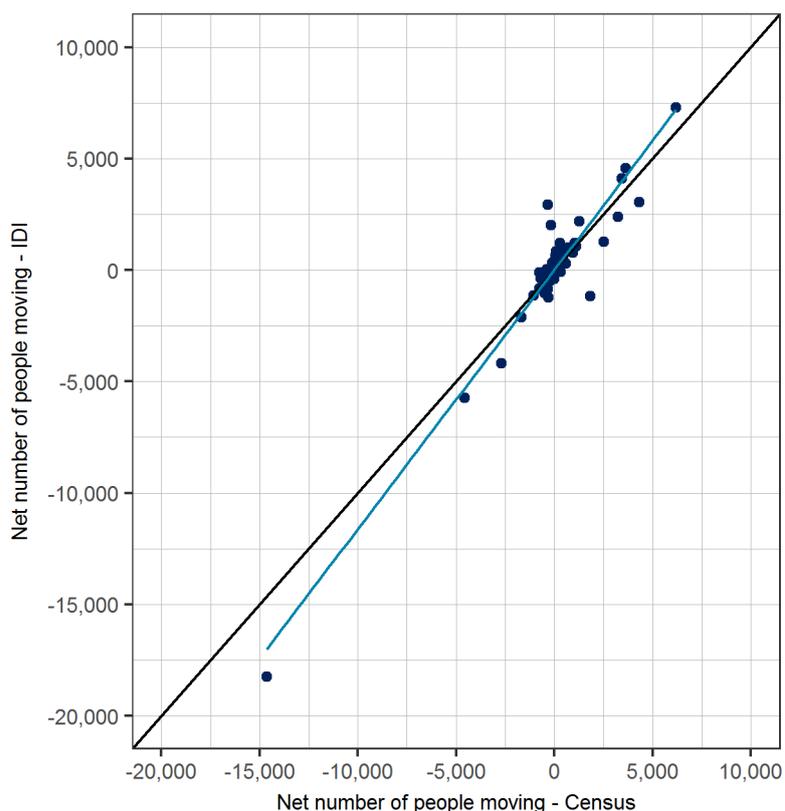


We may be interested in the total number of people moving into or out of a TA, but not necessarily the exact TAs people move to or from. Figure 12 shows total internal net migration figures between 2008 and 2013 for each of the 66 TAs included in the data.²⁷ In general the IDI and Census measures appear to relate well to each other. A steeper slope is again evident, as above, reflecting the larger flows measured using IDI. The correlation between the two measures is 0.956, showing a strong relationship between the two measures.

²⁶ The correlation coefficient is 0.970.

²⁷ As noted earlier the Chatham Islands Territory is excluded from the data.

Figure 12: Net internal migration flow comparison of Census and IDI-based figures – All TAs 2008 to 2013



Although the point at the bottom of the graph, representing Christchurch City, is most obviously different from the other points on the graph, this point is quite close to the regression line, indicating that the relationship between the Census and IDI measures is reasonably consistent with the overall pattern. Apart from Christchurch, the greatest divergence between IDI and Census net migration was for Wellington City, which had a net gain of 1,839 people according to Census, but a 1,173 net loss according to IDI. On the other side of the coin, in Whanganui District, Census showed a net loss of 336 people, while IDI showed a net gain of 2,919 people. Further work may be useful to better understand the source of these differences, which may derive from issues with the administrative data over these years. For example, NHI data for Whanganui does not appear to be captured in the IDI in the early part of this period, and that may have affected our internal migration estimates.

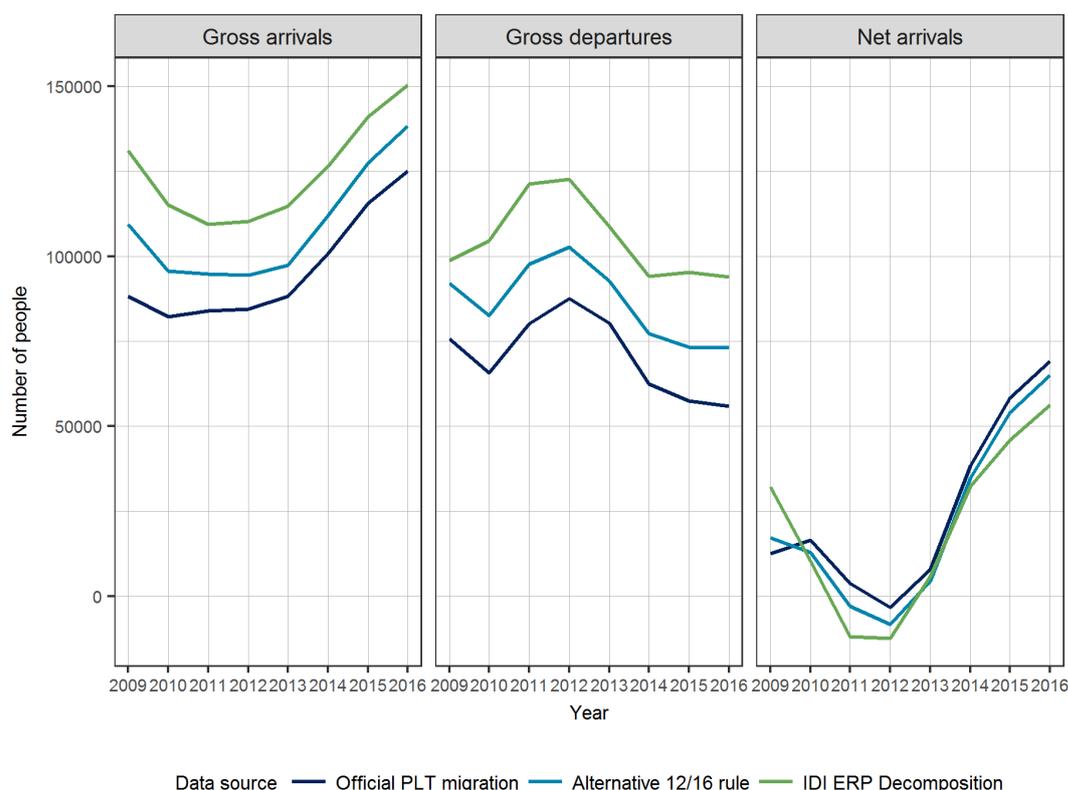
Differences between IDI and Census-based measures could be due to errors in either source. The quality of the IDI-based measure is dependent on address changes being identified quickly in government agency administrative systems. Where this doesn't happen moves could be identified in a period after they actually occur, or not identified at all. On the other hand, the Census measure is subject to undercounting due to non-response. This may introduce bias into the estimate and IDI may produce a more accurate estimate in some cases. Additionally, the quality of administrative location information held in the IDI has improved over time, and the IDI may produce a better measure of internal migration in more recent years. Finally, the IDI allows for internal migration measures to be produced more frequently than the Census, and for more recent periods.

International migration

The official statistical measure of international migration is based on identifying people as they cross the border into or out of New Zealand, and classifying them as permanent or long-term migrants largely based on the stated intentions in their arrival or departure cards. There are known issues with this measure, as illustrated in Stats NZ 2014. Stats NZ has subsequently produced a measure of migration using observed behaviour rather than intentions, as described in Stats NZ 2017a. The new measure tends to show higher number of both migrant arrivals and departures, while net arrivals may be higher or lower depending on the time period. Before 2009 the 12/16 rule produced higher net migration estimates, indicating that the official PLT measure was underestimating net migration – by more than 20,000 people in the year to December 2002. Since 2009 the 12/16 rule has resulted in slightly lower net migration figures than the official PLT figure.

Figure 13 compares these two figures across time since 2009, alongside the migration figures resulting from the decomposition of the IDI ERP outlined in this paper. We would expect some differences between our measure, using a modified version of the 12/16 rule, and Stats NZ's 12/16 measure. We use a different measure of residence, based on a fixed 16 month period between November of the previous year and February of the following year, while we also treat movement into or out of the ERP equally regardless of someone's history of residence (as discussed in section 2.4 above).

Figure 13: International migration comparison of official and IDI-based figures



The IDI ERP decomposition based measure discussed in this paper results in both higher gross arrivals and departures than either of the Stats NZ measures, but has resulted in lower net migration since 2009. In the year to June 2016 the official PLT net arrival figure was 69,000, compared to 65,000 using the 12/16 rule, and only 56,000 using our variation of the 12/16 rule.

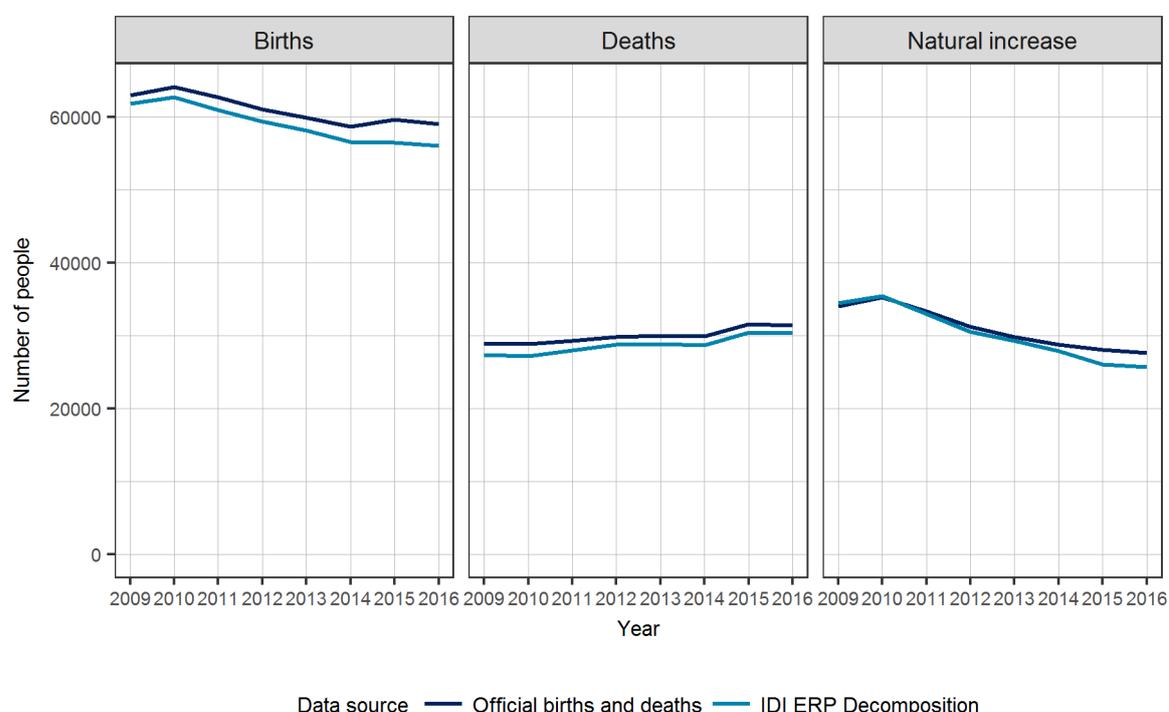
Natural increase

We would expect our measure of births, deaths, and natural increase (births minus deaths), to be similar to the official measure, as they are both derived from official birth and death records collected by the Department of Internal Affairs. One difference is that we use matched data from the IDI. While birth data is a key part of the construction of the IDI spine, death data is matched to the IDI through a subsequent process. Non-matches or matching error could have an influence on our results (particularly for deaths), although such effects are likely to be small as the match rate is high.

A second difference is that our birth and death figures are conditioned on the birth resulting in someone entering the IDI ERP in that year, or on the death resulting in someone exiting the IDI ERP in that year. In the case of births, a child may be excluded because they spend more than four months of the year outside the country after being born. In the case of deaths, someone who spent more than four months outside of New Zealand, and subsequently died in New Zealand, would not be included in the data. Deaths within a year of a birth are not included either in our birth or death figures, however this would not have an impact on natural increase.

Figure 14 compares official birth, death and natural increase figures²⁸ with our figures based on decompositions of changes in the IDI ERP. As expected, the figures are generally close over time. Both birth figures and death figures are slightly lower in our analysis, but natural increase is similar across most years. Official birth figures increased in 2015 and 2016, but this was not reflected in increases in births into the IDI ERP, resulting in some divergence in the estimate of natural increase. Further work is needed to better understand why this is the case.

Figure 14: Natural increase comparison of official and IDI-based figures



²⁸ Official birth, death and natural increase statistics were sourced from <https://www.stats.govt.nz/infoshare/>, are for New Zealand residents only, and are based on the date of registration.

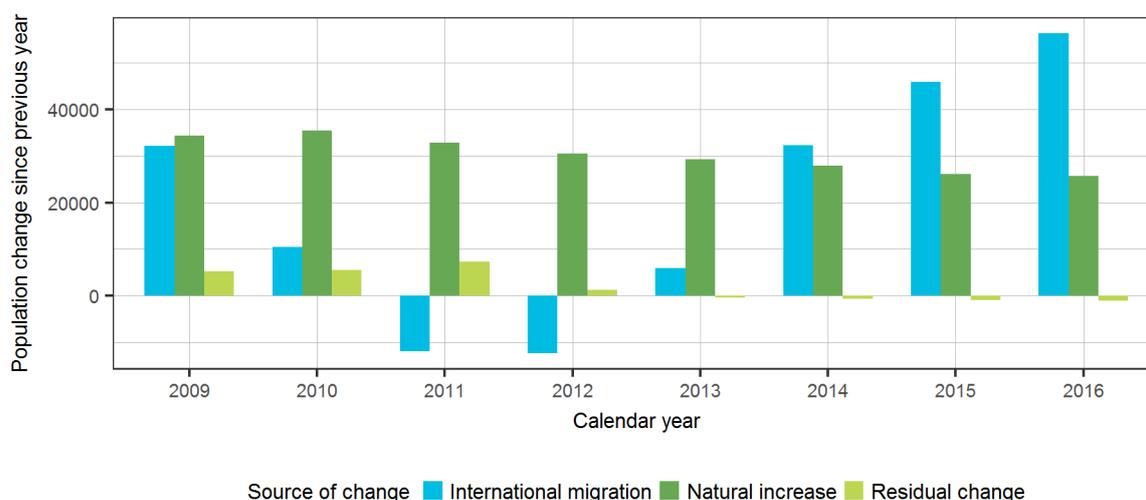
3 Key results

3.1 High level results

The new estimates of population change described in this report allow any period to be examined between the 2008 and 2016 calendar years. Differences between the populations of any two years can be described as being the result of natural change or international migration, and changes at a local area (TA) level can also be categorised as being the result of internal migration to and from other areas of the country. This section presents high level results for New Zealand across the entire period of interest, from 2008 to 2016, broken down by selected characteristics of interest. More detailed analysis can be undertaken using the Insights Population explorer and Population directions tools made available at <https://insights.apps.treasury.govt.nz/>.

Figure 15 shows the two main national-level population change components, international migration, and natural increase; as well as the small residual change that we were not able to account for. For each calendar year, the graph shows the change in the population since the previous year that was due to each component of population change. Population change due to international migration fluctuated considerably over time, while natural increase was relatively stable. The residual was larger in earlier years, as people were more difficult to locate in these years through administrative sources.

Figure 15: Total NZ decomposition of annual net population change, 2008 to 2016

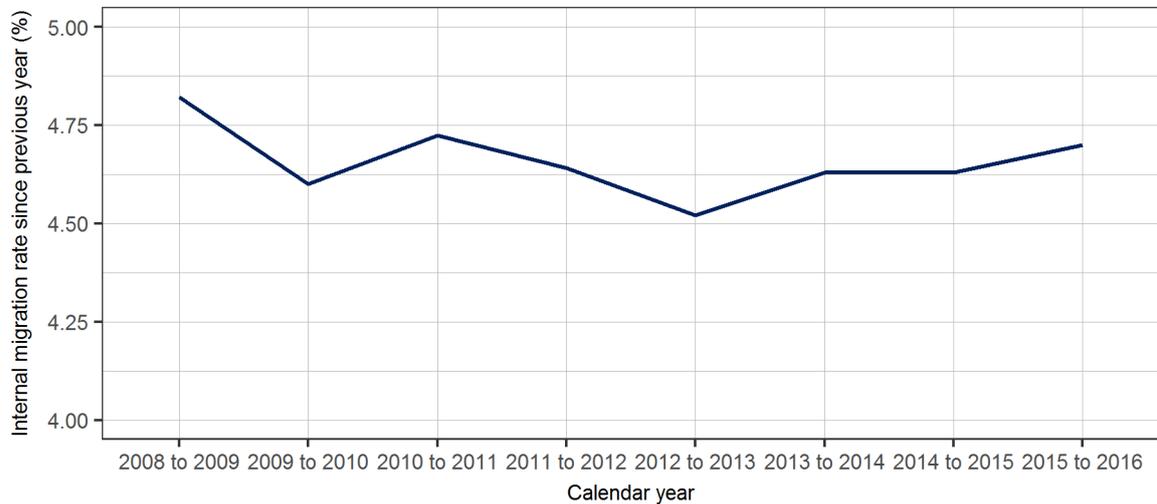


Migration within NZ

Net internal migration flows within New Zealand are equal to zero by construction. A person who moves into one area also moves out of another area. Similarly the total number of outflows across the country is equal to the number of inflows. As such, at a national level we are most interested in the total number of people moving in a particular year, and it makes sense to represent this as a percentage; either of the previous year's population, or of the population who were in New Zealand in both years.

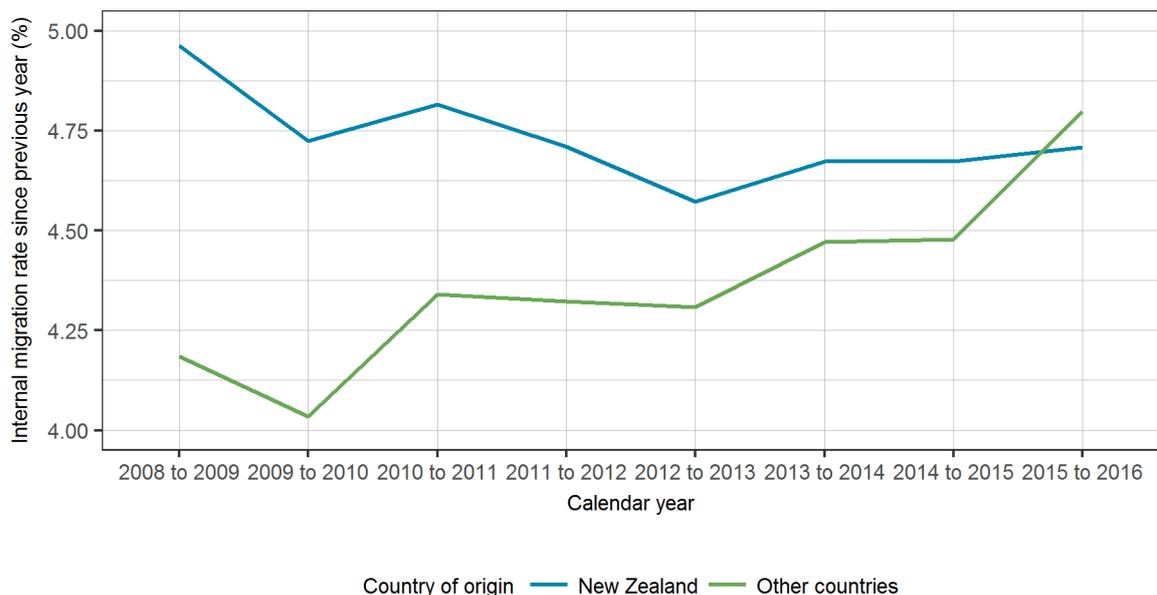
In Figure 16 below we present the internal migration rate over time, defined as the percentage of the New Zealand population in each year who move from one TA to another between that year and the following year. The internal migration rate is reasonably stable over time, ranging from 4.5% to 4.8%, with only small fluctuations from year to year.

Figure 16: Internal migration rate, 2008 to 2016



In Figure 17 we break this rate down by country of origin. See section 2.4 for a description of how country of origin is defined. While the New Zealand country of origin migration rate is reasonably stable over time, particularly over recent years, the non-New Zealand internal migration rate has risen considerably over time, passing the New Zealand rate for the first time in the 2016 calendar year. In that year recent migrants to New Zealand were more likely than earlier migrants or the New Zealand-born (the people we define as having a New Zealand country of origin) to move within New Zealand, albeit the rates were very similar – 4.8% and 4.7% respectively.

Figure 17: Internal migration rate by country of origin, 2008 to 2016



International Migration

Figure 18 shows net international migration by country of origin between 2008 and 2016. Although the number of New Zealanders²⁹ leaving the country has been larger than the number of New Zealanders arriving every year, there have only been relatively small negative net flows in recent years, dropping to a net loss of fewer than 4,500 New Zealanders from 2015 to 2016, down from a loss of 40,500 between 2011 and 2012. The largest source of migrants to New Zealand has been India in every year since 2009, although China also been a large source of migrants in recent years.

Figure 18: Total NZ net population change due to international migration by country of origin, 2008 to 2016 calendar years

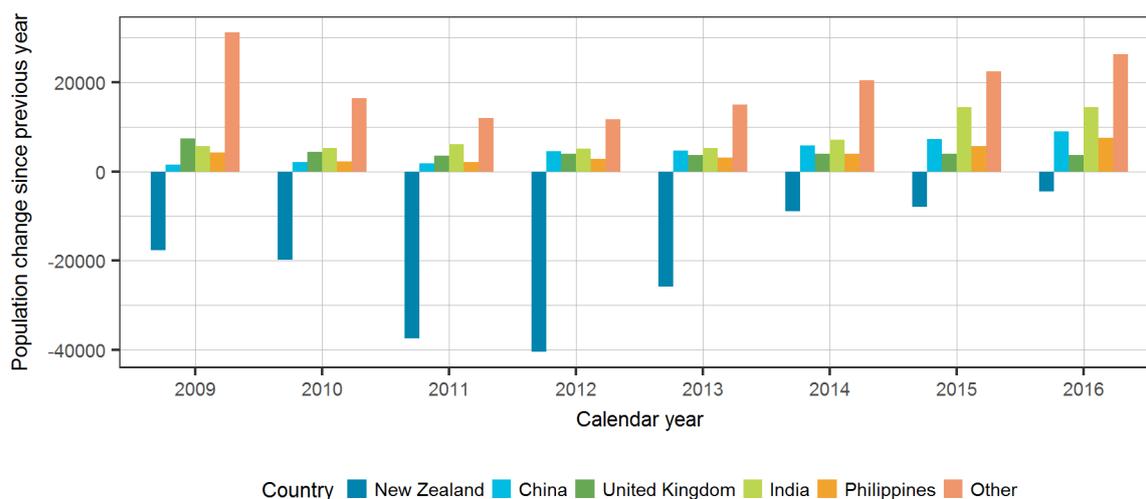


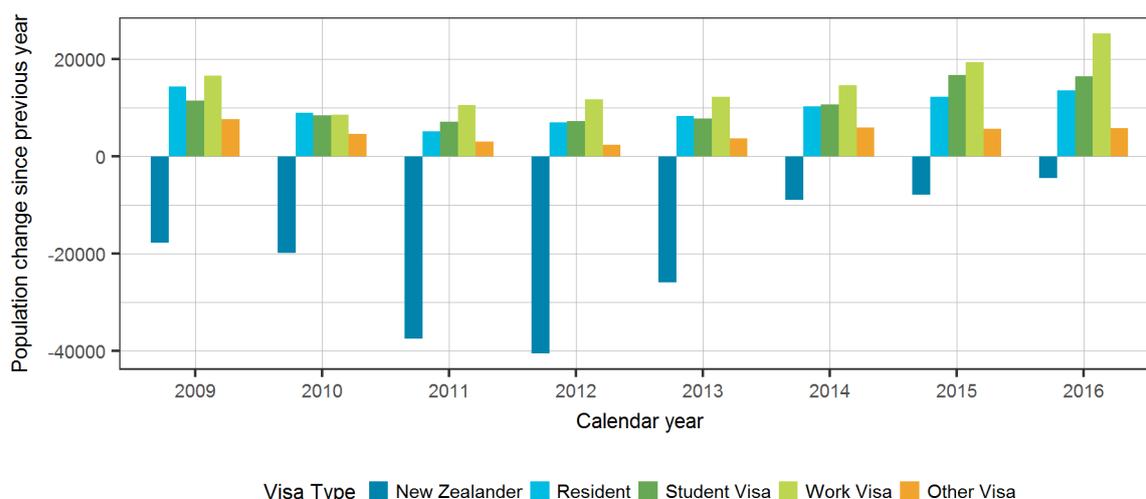
Figure 19 breaks down international migrant arrivals by visa category each year since 2008. Migrants enter New Zealand on a range of visas, which can be broadly captured under the categories of: residence, where a migrant is granted a permanent right to reside in New Zealand; student, where a person is granted the temporary right to study in New Zealand; and work, where a person is granted the temporary right to work in New Zealand. The other visa category included in Figure 19 includes primarily: Australian citizens, who have a permanent right to reside in New Zealand without the need for a residence visa; and visitors, who generally do not have the right to work and only have limited rights to study.

Many migrants to New Zealand first arrive with a student or work visa, and over time progress to work and residence³⁰. As Figure 19 focuses on year-to-year changes in the population, many migrants are identified as being on temporary visas, while a focus on migration over longer periods will result in relatively more people being identified as residents. In the early period there was a fairly even mix between the three main types of visa, however work visas made by far the largest contribution to the growth in the IDI ERP between 2015 and 2016, with a net inflow of 25,000, compared to a net inflow of 13,000 residents, and 16,000 international students.

²⁹ In the context of this paper New Zealanders is taken to mean people with a New Zealand 'country of origin' using the definition outlined in Section 2.4. To be classified in this way a person must have been resident in New Zealand for at least 10 years, but they may not have been born here.

³⁰ Many migrants also arrive on student or work (especially working holiday) visas, but do not stay long enough to ever meet the 12/16 criteria and are therefore never considered to have become part of the IDI ERP. These shorter-term migrants are not included in our statistics.

Figure 19: Total NZ net population change due to international migration by visa type, 2008 to 2016 calendar years



In sections 3.2 to 3.4 below, we present three short case studies looking at population change in specific territorial authorities: Christchurch, which experienced significant population change over our period of interest as a result of large earthquakes in 2010 and 2011; Queenstown, which has experienced rapid tourism-fuelled growth in recent years; and Auckland, New Zealand’s largest city.

Despite differences in the definition of residence, time periods covered, and measurement approach, the analysis presented in the three case studies shows a broadly similar picture at an overall level to that shown in official subnational population estimates.

- In Christchurch, both sources show similar patterns of population decline and recovery, with the population in 2016 being very similar to the 2010 pre-earthquake level.
- In Auckland, both sources show consistent year-on-year growth, with almost identical figures between 2008 and 2013, however the extent of that growth differs in recent years. According to official estimates the ERP in Auckland grew from 1.49 million to 1.61 million from 2013 to 2016. Our estimates show a slower growth over the period, from 1.50 million to 1.58 million. Net migration (combining internal and international migration) totalled almost 80,000 over those three years according to Stats NZ estimates. Our estimates were only half that, at around 40,000 (made up of a 65,000 increase due to international migration and a 15,000 loss due to migration within NZ). Further work is required to better understand these differences. They may relate to definitional differences, such as in the way usual residence is defined, or may relate to measurement differences, for example in the way location is determined. It is unclear whether differences are driven by the internal or international migration components, or a mixture of the two.³¹

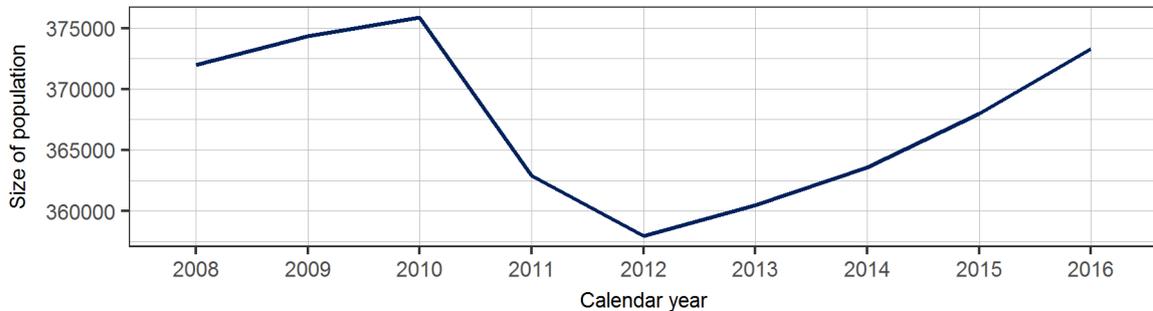
³¹ It is also worth noting that official population estimates are likely to be less reliable as time elapses after the most recent Census. Once a new Census is undertaken these estimates may be revised. For example, provisional estimates of the New Zealand and Auckland resident populations as at June 2013 were 4.47 million and 1.53 million respectively. Following the 2013 Census, the New Zealand estimated resident population was revised downward by 29,000 and the Auckland population revised downward by 36,000 to 4.44 million and 1.49 million respectively.

- The Queenstown-Lakes district population grew from 26,000 to 35,000 between 2008 and 2016 according to official sources, with growth accelerating in recent years (increasing by 2,300 in the year to June 2016). These figures are almost identical to our estimates; an increase from 25,000 to 35,000 over the same period, and an increase of 2,400 between 2015 and 2016.

3.2 Case study 1 – Christchurch

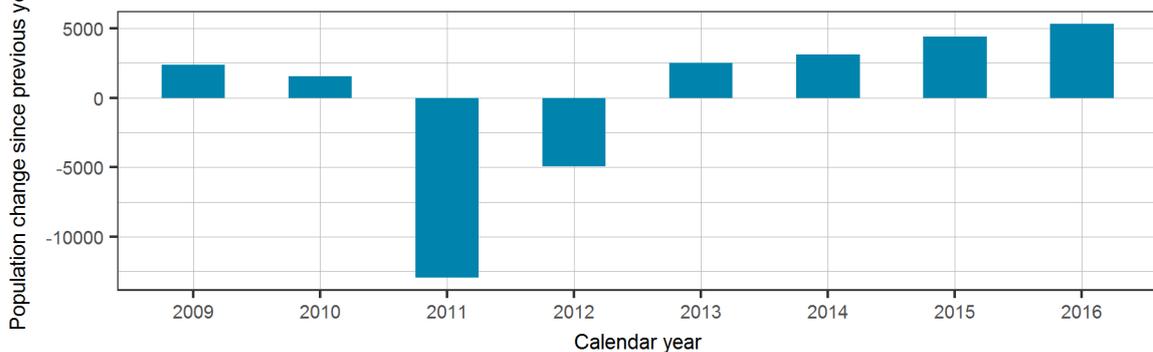
The population of Christchurch City fell in 2011 and 2012, following the earthquakes in September 2010 and February 2011, before re-bounding in subsequent years. In 2016 the population was 373,000, slightly lower than the 2010 peak of 376,000.

Christchurch City Estimated Resident Population, 2008 to 2016



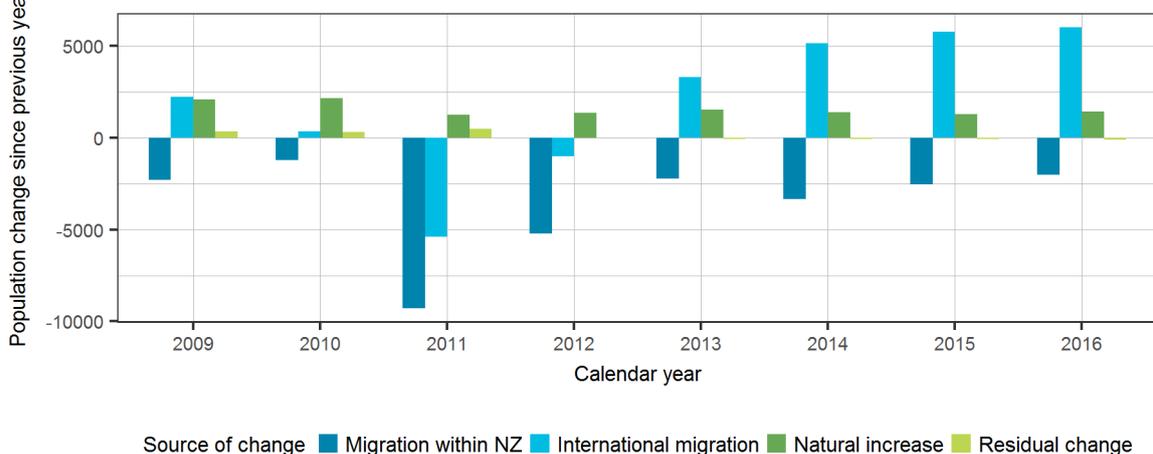
The population of Christchurch City increased by around 2,500 to 5,000 people per year in each of the years since 2012, after falling by a total of 18,000 in the two years to 2012.

Christchurch City IDI estimated resident population annual net change 2008 to 2016

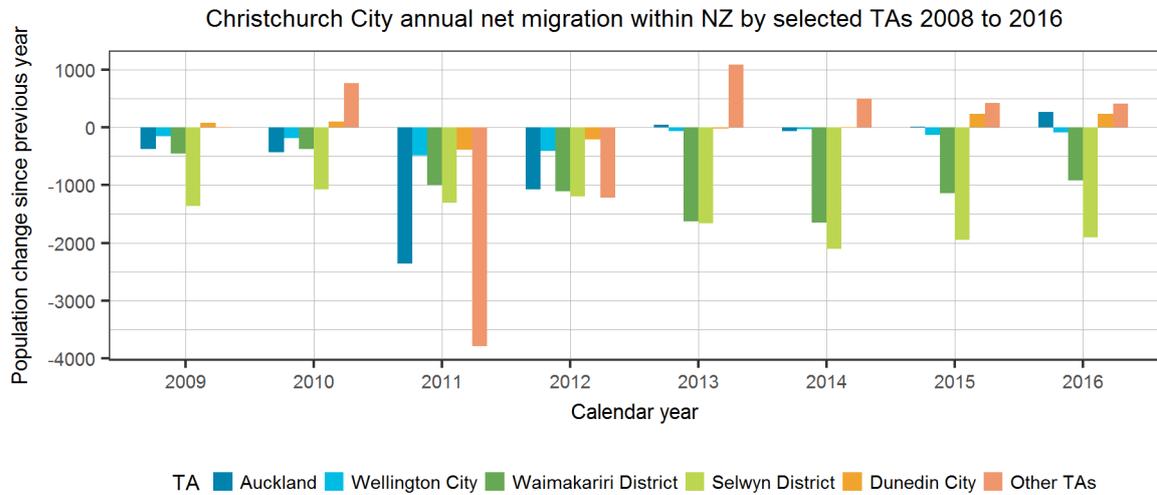


The fall in the population from 2010 to 2011 was due to people moving from Christchurch to other parts of New Zealand, and to a lesser degree, to people moving overseas. From 2011 to 2012 almost all of the decrease was due to people moving within New Zealand. Since 2012 increases in the population have been driven by people arriving from overseas, more than offsetting the continued net movement of people moving to other parts of the country.

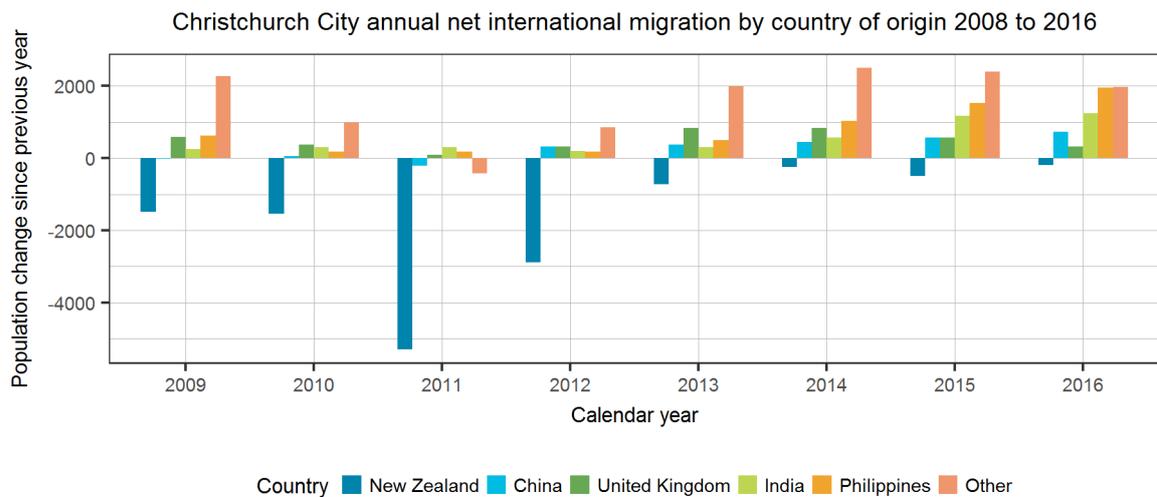
Christchurch City decomposition of annual net population change 2008 to 2016



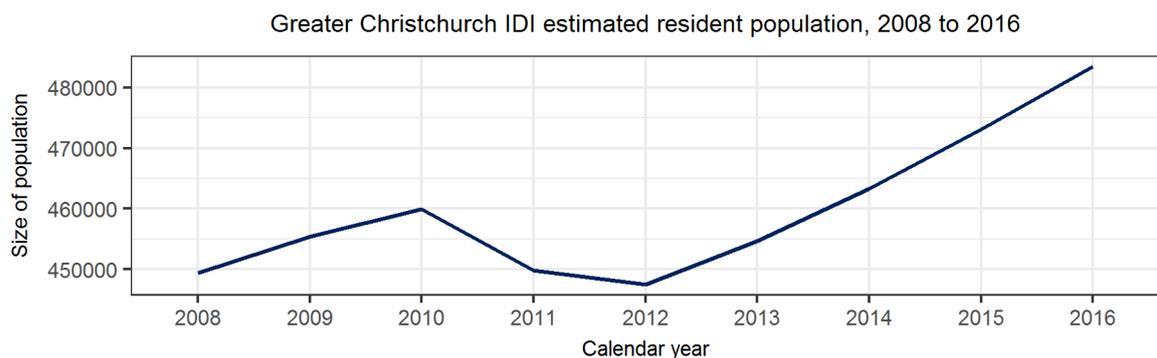
From 2010 to 2012, people leaving Christchurch moved to a number of different areas of the country, but particularly Auckland and the adjacent Selwyn and Waimakariri districts. Since 2012 almost all of the movement has been to these latter two areas, while there has been a flow into Christchurch City from other areas.



The net outflow of people from Christchurch to other countries from 2010 to 2012 was largely driven by a net outflows of New Zealanders. Since 2012 almost all of the growth was due to flows of migrants from other countries, especially the Philippines and India.



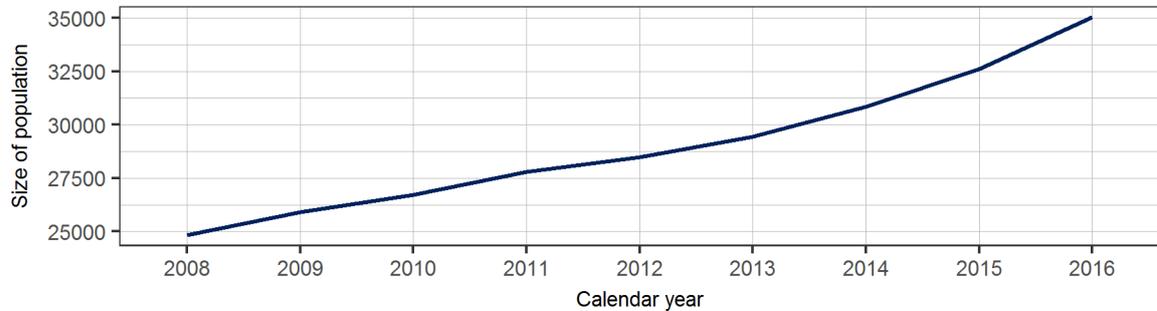
Although the population of Christchurch City was still smaller in 2016 than 2010, the population of the Greater Christchurch area, encompassing Christchurch City and the adjacent Selwyn and Waimakariri Districts, was considerably larger. The population of this broader area grew from 460,000 to 483,000 since 2010, after falling to 447,000 in 2012, following the 2010/11 earthquakes.



3.3 Case study 2 – Queenstown-Lakes District

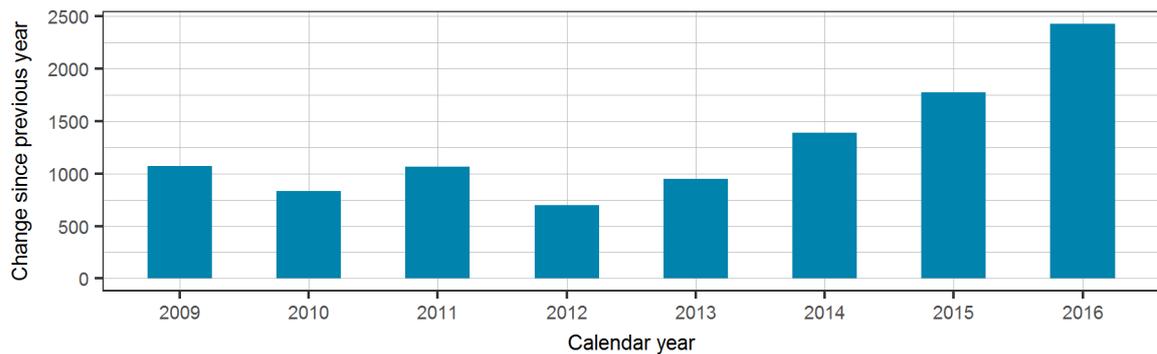
The Queenstown-Lakes district includes the tourism resort towns of Queenstown and Wanaka, as well as a number of smaller population centres. The population of Queenstown-Lakes increased every year from 2008 to 2016, with a total increase of more than 10,000 people or 41 percent over the period.

Queenstown-Lakes District IDI estimated resident population, 2008 to 2016



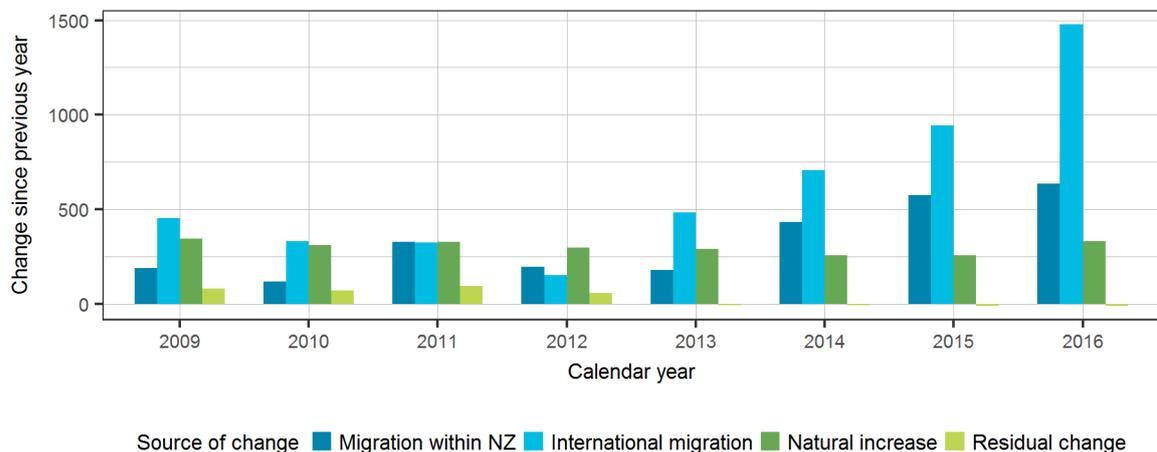
Year-on-year net changes in the population have been particularly high in recent years, and increasing each year, from an increase of fewer than 700 people from 2011 to 2012, up to almost 2,500 from 2015 to 2016.

Queenstown-Lakes District Estimated Resident Population annual net change 2008 to 2016



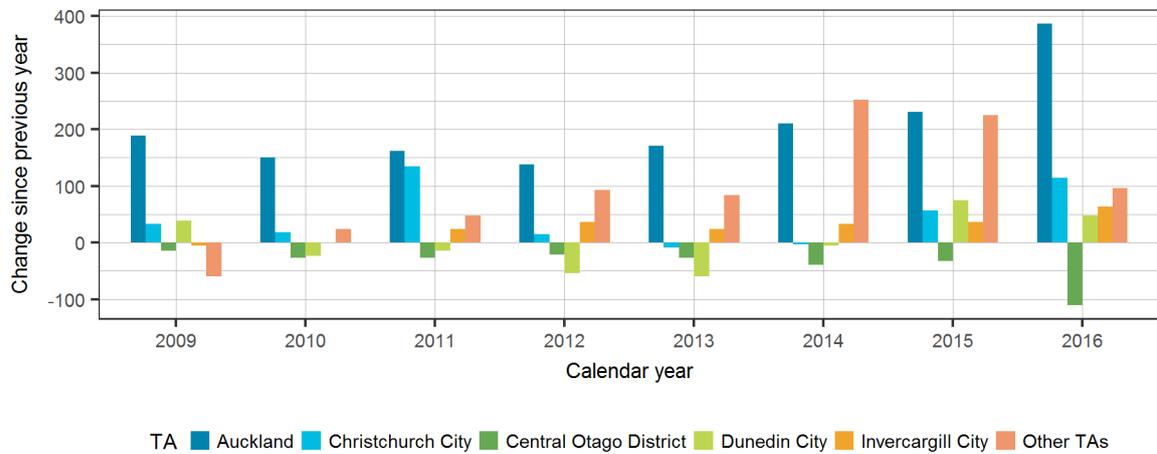
Increases in the Queenstown-Lakes population between 2008 and 2012 were driven by approximately equal parts migration from other parts of NZ, international migration and natural increase. In recent years, the accelerating population growth has been driven largely by international migration flows, and to a lesser degree by increasing flows from other parts of the country.

Queenstown-Lakes District decomposition of annual net population change 2008 to 2016



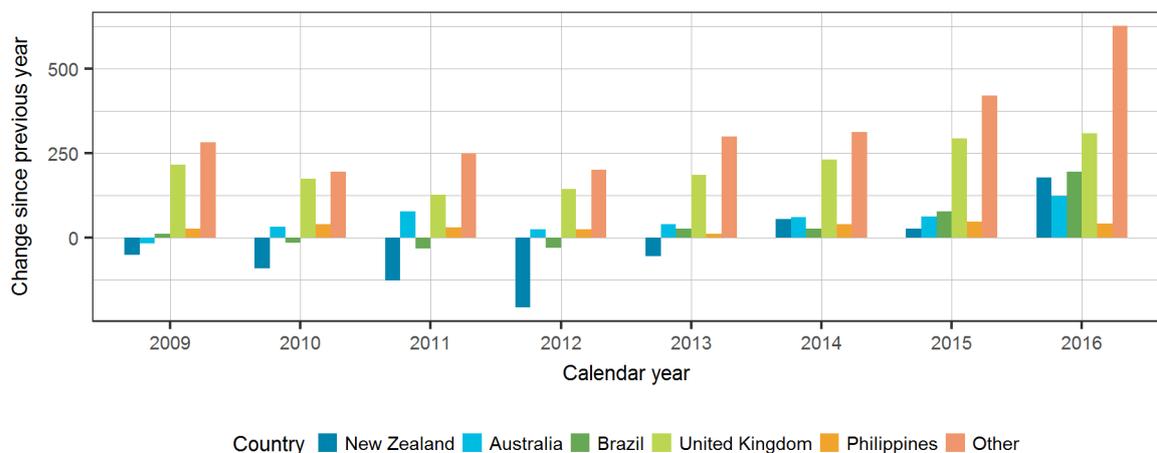
In most years, the largest flow into Queenstown-Lakes from other areas of New Zealand has been from Auckland, although this was almost matched by flows from Christchurch between 2010 and 2011 following the earthquakes. The only net flows out of Queenstown of any significance were flows to Central Otago from 2015 to 2016.

Queenstown-Lakes District annual net migration within NZ by selected TAs 2008 to 2016



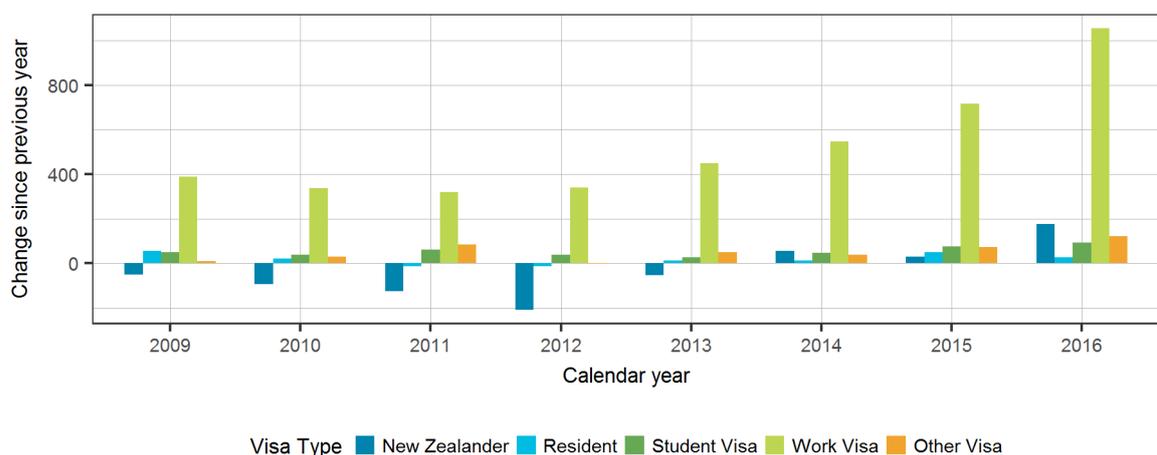
As discussed above, most of the growth in the Queenstown-Lakes population in recent years was driven by international migration. The United Kingdom remains the largest single source country, although Brazil emerged as a key source in 2016.

Queenstown-Lakes District annual net international migration by country of origin 2008 to 2016



Most non-New Zealanders arriving in Queenstown were temporary work visa holders.

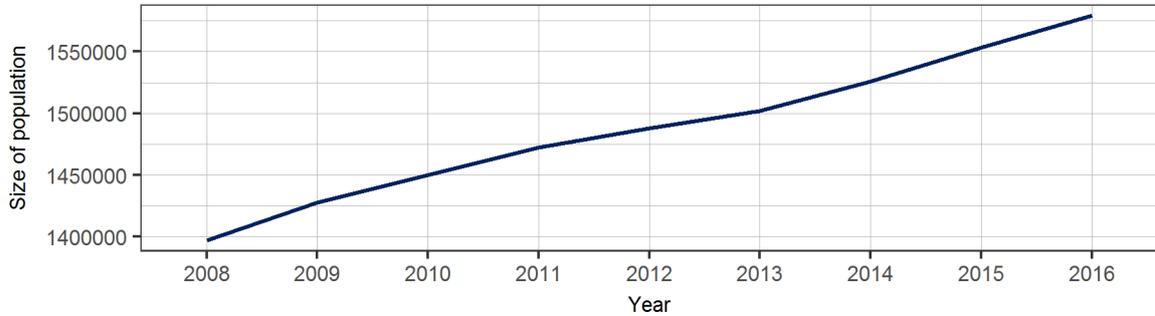
Queenstown-Lakes District annual net international migration by visa type 2008 to 2016



3.4 Case study 3 – Auckland

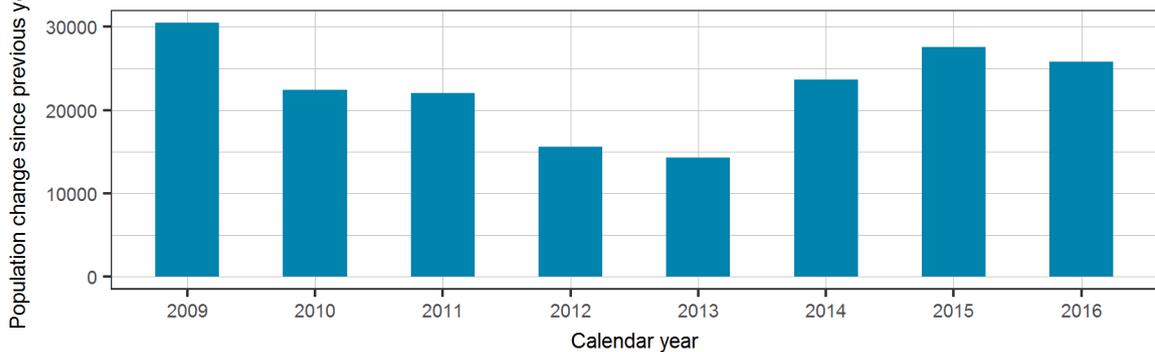
The Auckland territorial authority includes over a third of the New Zealand IDI ERP, and has grown considerably since 2008 – from 1.40 million in 2008 to 1.58 million in 2016, an increase of 180,000 or 13 percent over the period.

Auckland IDI estimated resident population, 2008 to 2016



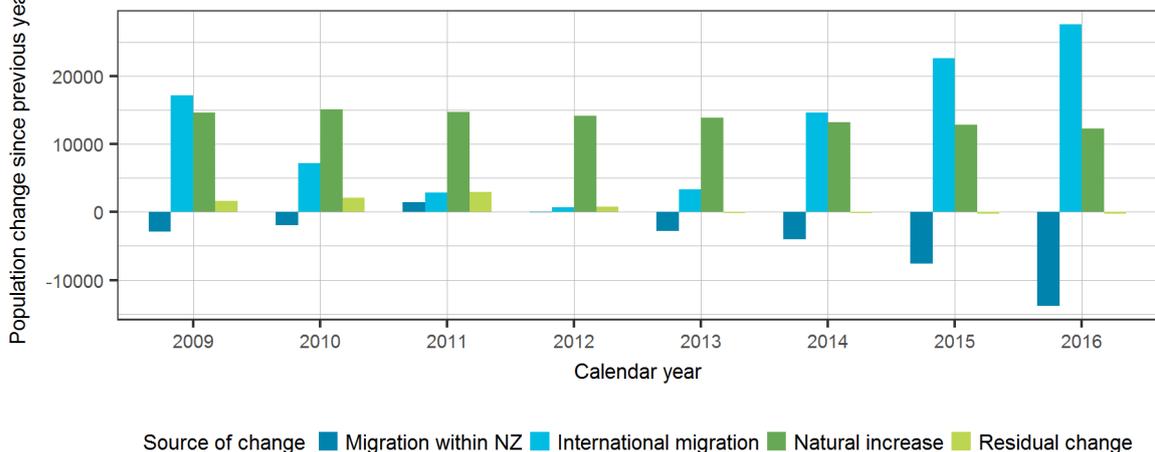
Year-on-year net changes in the population have been fairly consistent across the entire period, with lows of around 15,000 per year in the 2012 and 2013 calendar years. The population has increased by at least 20,000 in each of the other years.

Auckland Estimated Resident Population annual net change 2008 to 2016

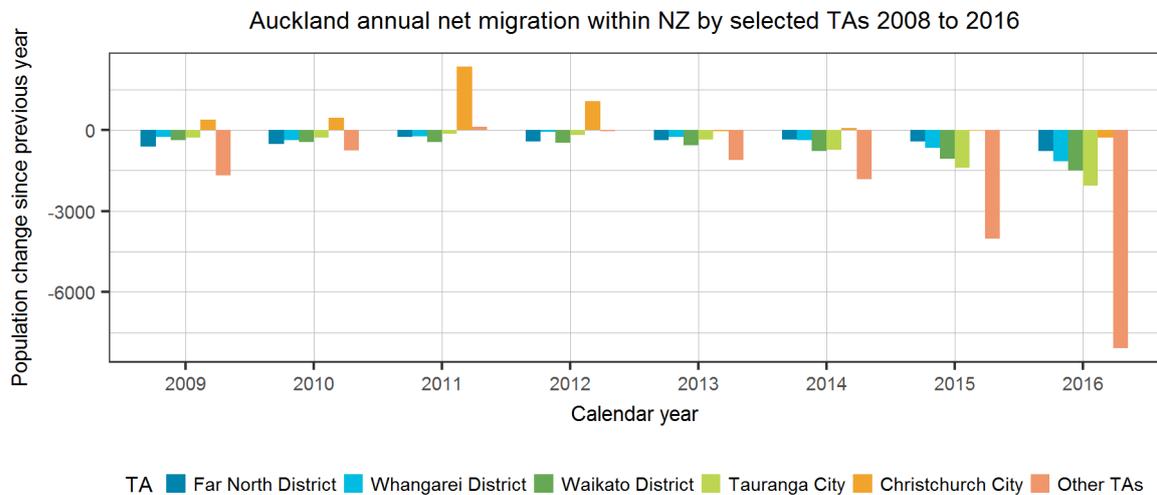


Natural increase has been a fairly constant source of increase in the Auckland population, while international migration has fluctuated, contributing fewer than a thousand additional people to the Auckland population between 2011 and 2012, before rising to almost 28,000 from 2015 to 2016. As international migration has grown in recent years internal migration away from Auckland to other parts of New Zealand has also grown. From 2015 to 2016 Auckland lost almost 14,000 residents to other TAs.

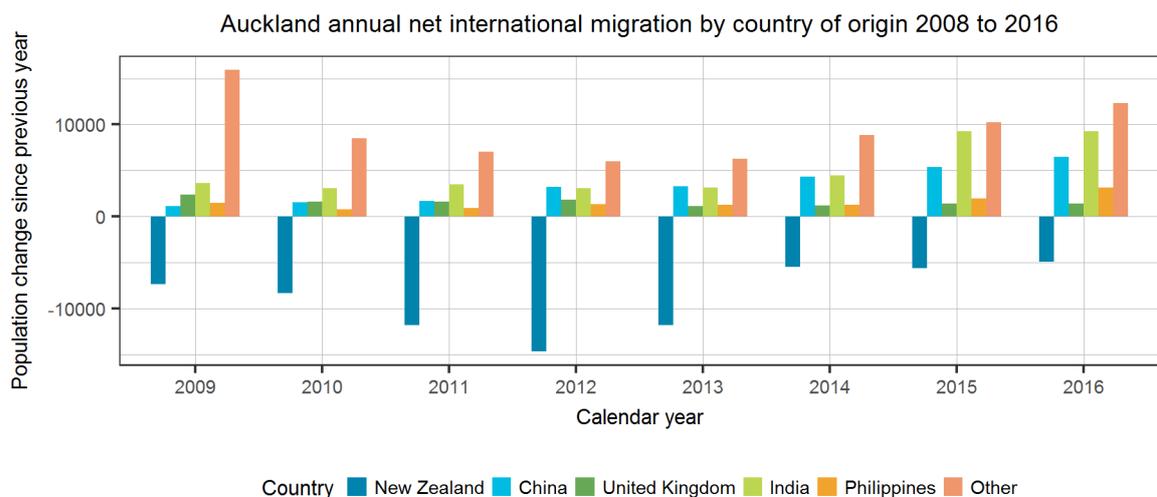
Auckland decomposition of annual net population change 2008 to 2016



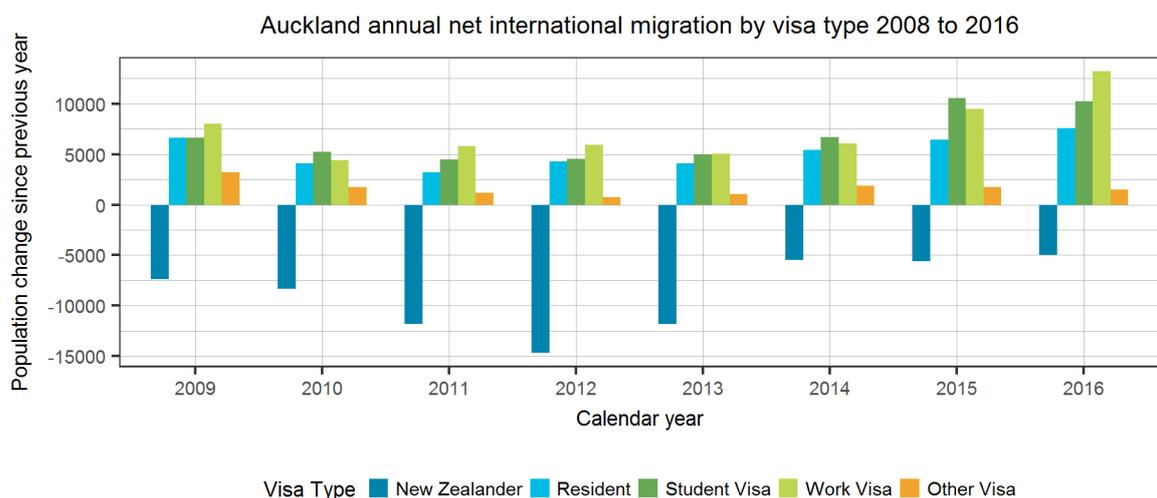
The main movement out of Auckland in recent years was to nearby TAs (the Far North, Whangarei and Waikato Districts, and Tauranga City), but large numbers also moved to other TAs around the country. The only significant net flow into Auckland was from Christchurch following the 2010 and 2011 earthquakes.



There has been a consistent net flow of New Zealanders leaving Auckland to go overseas, albeit at a reduced rate in recent years. Non-New Zealand migration into Auckland has been dominated by migrants from China and India in recent years.



Migrants arriving in Auckland held a mix of work, student and residence visas.



4 Next steps

This report outlines preliminary work to describe population change in New Zealand using integrated administrative data. It highlights and confirms the enormous potential for local area population estimation using administrative data sources. This work could be further developed in a number of different ways.

Future work could look at using more-detailed geographic classifications, particularly in the large Auckland metropolitan area. It could also look at describing population change using a broader range of data that is readily available in the IDI. Analysis could focus, for example, on changes in the characteristics of an area's workforce, or could look at the drivers of changing utilisation of government services, such as healthcare or education.

Further work could also be done to improve the allocation of people to residential locations developed in this paper, and this could be further tested and refined using the 2018 Census, once that becomes available.

Towards the end of 2018 this data could be extended to include the 2017 year. The code used to develop and describe this data will be made publically available and as such can be picked up and further developed by any researchers with an interest in the work.

The data described in this paper could also be used to undertake analytical work that looks at estimating the impact of different types of population change, for example on the housing market. Research could test the association between geographical and temporal variation in population change and a number of different outcome measures of interest.

References

- Black, A. (2016). *The IDI prototype spine's creation and coverage*. (Statistics New Zealand Working Paper No 16–03). Retrieved from www.stats.govt.nz.
- Breiman, L. (1996). *Bagging predictors*. *Machine Learning*. 24 (2): 123–140.
- McLeod, K. and Tumen, S. (2017). *Insights - Informing policies and services for at-risk children and youth*. Treasury Analytical Paper 17/02.
- Stats NZ (2011). *Evaluation of alternative data sources for population estimates*. Retrieved from www.stats.govt.nz.
- Stats NZ (2014). *Alternative methods for measuring permanent and long-term migration*. Retrieved from www.stats.govt.nz.
- Stats NZ (2017a). *Defining migrants using travel histories and the '12/16-month rule'*. Retrieved from www.stats.govt.nz.
- Stats NZ (2017b). *Experimental population estimates from linked administrative data: 2017 release*. Retrieved from www.stats.govt.nz.
- Wolpert, D. (1992). *Stacked generalization*. *Neural Networks*. 5(2): 241-259.

Appendix

Table 10: Sources of location data in the IDI

Data source	Description	Period coverage*	No. of records (million)
2013 Census (current and 5 years ago)	Location at March 2013 and March 2008. Some of the 2013 population were not in NZ in 2008 while others did not give a useable address.	2008 and 2013	3.9 / 3.2
ACC claims	Client address for ACC claims	1994 to 2016	17.8
Arrival cards	Addresses from Arrival Cards are coded to territorial authority.	1998 to 2017	1.9
Departure cards	Addresses from Departure Cards are coded to territorial authority.	1998 to 2017	3
Drivers' license	Drivers' Licence register. Addresses on the register have been gradually increasing over time, however the register does not contain historical records.	1954 to 2017	4.3
Employer location	Derived from the identified location of the business a salary and wage earner works for.	2002 to 2017	24.8
HES	Household Economic Survey	2007 to 2015	0.08
HLFS	Household Labour Force Survey	2007 to 2017	1.6
Housing New Zealand	Addresses from new applications, transfers and a register snapshot.	2003 to 2015	1.8
IR	Inland Revenue tax registrations address	2001 to 2017	40.6
MOE school enrolment	Address from school enrolment.	2007 to 2016	0.9
MOE school location	Location of school attended.	2007 to 2016	3.2
MOE tertiary provider location	TA location of tertiary providers, excluding those providing courses in multiple areas.	2003 to 2016	8.4
Motor vehicle license	Motor vehicle license register. Addresses on the register have been gradually increasing over time, however the register does not contain historical records.	1911 to 2017	3.7
MSD residential / postal	Ministry of Social Development – National Superannuation and Benefit system addresses	1991 to 2015	13.2 / 1.17
NHI register	Addresses from National Health Index Register	2004 to 2017	24.8
PHO register	Address from registration with Primary Health Organisations	2003 to 2017	18.5

* The coverage period reflects the time over which the majority of location information has been collected. In some cases there have been significant improvements in the collection across the period.